



## DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITE DE COOPERATION EN MATIERE DE BREVETS (PCT)

<p>(51) Classification internationale des brevets <sup>6</sup> :</p> <p><b>C12N 15/74, 15/31, C07K 14/35, A61K 48/00, 39/04, C07K 19/00, C12Q 1/68, C07K 16/12, G01N 33/50, 33/53 // C12N 15/52, 15/65</b></p>	<p><b>A2</b></p>	<p>(11) Numéro de publication internationale: <b>WO 99/09186</b></p> <p>(43) Date de publication internationale: 25 février 1999 (25.02.99)</p>						
<p>(21) Numéro de la demande internationale: PCT/FR98/01813</p> <p>(22) Date de dépôt international: 14 août 1998 (14.08.98)</p> <p>(30) Données relatives à la priorité:</p> <table border="0"> <tr> <td>97/10404</td> <td>14 août 1997 (14.08.97)</td> <td>FR</td> </tr> <tr> <td>97/11325</td> <td>11 septembre 1997 (11.09.97)</td> <td>FR</td> </tr> </table>		97/10404	14 août 1997 (14.08.97)	FR	97/11325	11 septembre 1997 (11.09.97)	FR	<p>(74) Mandataires: MARTIN, Jean-Jacques etc.; Cabinet Regimbeau, 26, avenue Kléber, F-75116 Paris (FR).</p> <p>(81) Etats désignés: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, brevet ARIPO (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), brevet eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet européen (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), brevet OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p>
97/10404	14 août 1997 (14.08.97)	FR						
97/11325	11 septembre 1997 (11.09.97)	FR						
<p>(71) Déposant (pour tous les Etats désignés sauf US): INSTITUT PASTEUR [FR/FR]; 28, rue du Docteur Roux, F-75015 Paris (FR).</p> <p>(72) Inventeurs; et</p> <p>(75) Inventeurs/Déposants (US seulement): GICQUEL, Brigitte [FR/FR]; 8, rue Daguerre, F-75014 Paris (FR). PORTNOÏ, Denis [FR/FR]; 7, rue Simon LeFranc, F-75004 Paris (FR). LIM, Eng-Mong [KH/FR]; 20, rue Georges Pitard, F-75015 Paris (FR). PELICIC, Vladimir [FR/FR]; 28, rue de Chateaudun, F-75009 Paris (FR). GUIGUENO, Agnès [FR/FR]; 26-28, rue Gambetta, F-62026 Arras (FR). GOGUET DE LA SALMONIERE, Yves [FR/FR]; 30, rue Lourmel, F-75015 Paris (FR).</p>		<p>Publiée</p> <p><i>Sans rapport de recherche internationale, sera republiée dès réception de ce rapport.</i></p> <p><i>Avec une indication relative à du matériel biologique déposé, fournie selon la règle 13bis, séparément, et non avec la description.</i></p>						
<p>(54) Title: POLYPEPTIDE NUCLEIC SEQUENCES EXPORTED FROM MYCOBACTERIA, VECTORS COMPRISING SAME AND USES FOR DIAGNOSING AND PREVENTING TUBERCULOSIS</p>								
<p>(54) Titre: SEQUENCES NUCLEIQUES DE POLYPEPTIDES EXPORTES DE MYCOBACTERIES, VECTEURS LES COMPRENANT ET APPLICATIONS AU DIAGNOSTIC ET A LA PREVENTION DE LA TUBERCULOSE</p>								
<p>(57) Abstract</p> <p>The invention concerns recombinant vectors replicated in mycobacteria, a set of sequences coding for exported polypeptides detected by fusion with alkaline phosphatase, in particular one polypeptide, called DP428, of about 12 kD corresponding to an exported protein found in mycobacteria belonging to the <i>Mycobacterium tuberculosis</i> complex. The invention also concerns methods and kits for detecting in vitro the presence of a mycobacterium and in particular a mycobacterium belonging to the <i>Mycobacterium tuberculosis</i> complex in a biological sample using said polypeptides, their fragments or polynucleotides coding for the latter. The invention also concerns immunogenic or vaccine compositions for preventing and/or treating infections caused by mycobacteria and in particular a mycobacterium belonging to said complex, particularly tuberculosis.</p>								
<p>(57) Abrégé</p> <p>L'invention a pour objet des vecteurs recombinants se répliquant chez les mycobactéries, un ensemble de séquences codant pour des polypeptides exportés détectés par des fusions avec la phosphatase alcaline, notamment un polypeptide, dénommé DP428, d'environ 12kD correspondant à une protéine exportée retrouvée dans les mycobactéries appartenant au complexe de <i>Mycobacterium tuberculosis</i>. L'invention concerne également des procédés et des kits de détection <i>in vitro</i> de la présence d'une mycobactérie et en particulier une mycobactérie appartenant au complexe de <i>Mycobacterium tuberculosis</i> dans un échantillon biologique utilisant lesdits polypeptides, leurs fragments ou des polynucleotides codant pour ces derniers. L'invention vise des compositions immunogènes ou vaccins pour la prévention et/ou le traitement d'infections provoquées par des mycobactéries et en particulier une mycobactérie appartenant audit complexe, en particulier la tuberculose.</p>								

# UNIQUEMENT A TITRE D'INFORMATION

Codes utilisés pour identifier les Etats parties au PCT, sur les pages de couverture des brochures publiant des demandes internationales en vertu du PCT.

AL	Albanie	ES	Espagne	LS	Lesotho	SI	Slovenie
AM	Arménie	FI	Finlande	LT	Lituanie	SK	Slovaquie
AT	Autriche	FR	France	LU	Luxembourg	SN	Sénégal
AU	Australie	GA	Gabon	LV	Lettonie	SZ	Swaziland
AZ	Azerbaïdjan	GB	Royaume-Uni	MC	Monaco	TD	Tchad
BA	Bosnie-Herzégovine	GE	Géorgie	MD	République de Moldova	TG	Togo
BB	Barbade	GH	Ghana	MG	Madagascar	TJ	Tadjikistan
BE	Belgique	GN	Guinée	MK	Ex République yougoslave de Macédoine	TM	Turkménistan
BF	Burkina Faso	GR	Grèce	ML	Mali	TR	Turquie
BG	Bulgarie	HU	Hongrie	MN	Mongolie	TT	Trinité et Tobago
BJ	Bénin	IE	Irlande	MR	Mauritanie	UA	Ukraine
BR	Bresil	IL	Israël	MW	Malawi	UG	Ouganda
BY	Belarus	IS	Islande	MX	Mexique	US	Etats-Unis d'Amérique
CA	Canada	IT	Italie	NE	Niger	UZ	Ouzbékistan
CE	République centrafricaine	JP	Japon	NL	Pays-Bas	VS	Viet Nam
CG	Congo	KE	Kenya	NO	Norvège	YU	Yougoslavie
CH	Suisse	KG	Kirghizistan				

FR  
11

FR  
11

FR  
11

Séquences nucléiques de polypeptides exportés de mycobactéries, vecteurs les comprenant et applications au diagnostic et à la prévention de la tuberculose.

5 L'invention a pour objet de nouveaux vecteurs recombinants de criblage, de clonage et/ou d'expression se répliquant chez les mycobactéries. Elle a également pour objet un ensemble de séquences codant pour des polypeptides exportés détectés par des fusions avec la phosphatase  
10 alcaline et dont l'expression est régulée (induite ou réprimée) ou constitutive lors de l'ingestion des mycobactéries par les macrophages. L'invention concerne également un polypeptide, dénommé DP428, d'environ 12kD correspondant à une protéine exportée retrouvée dans les  
15 mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*. L'invention vise aussi un polynucléotide comprenant une séquence codant pour ce polypeptide. Elle concerne également l'utilisation du polypeptide ou de fragments de celui-ci et des polynucléotides codant pour  
20 ces derniers (ou encore les polynucléotides complémentaires à ces derniers) pour la réalisation de moyens de détection in vitro, ou in vivo de la présence d'une mycobactérie appartenant au complexe de *Mycobacterium tuberculosis* dans un échantillon biologique ou pour la détection de réactions  
25 de l'hôte infecté par ces espèces bactériennes. L'invention vise enfin l'utilisation du polypeptide ou de fragments de celui-ci ainsi que des polynucléotides codant pour ces derniers en tant que moyens destinés à la préparation d'une composition immunogène, susceptible d'induire une réponse  
30 immunitaire dirigée contre les mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*, ou d'une

La présente invention a aussi pour but d'utiliser ces séquences (polypeptidiques et polynucléotidiques) comme  
5 cible pour la recherche de nouveaux inhibiteurs de la croissance et de la multiplication des mycobactéries et de leur maintien chez l'hôte, ses inhibiteurs pouvant servir d'antibiotiques.

10 Le genre *Mycobacterium*, qui comprend au moins 56 espèces différentes, inclut des pathogènes humains majeurs tels que *M. leprae* et *M. tuberculosis*, les agents responsables de la lèpre et de la tuberculose, qui restent des problèmes graves de santé publique dans le monde  
15 entier.

La tuberculose continue d'être un problème de santé publique dans le monde. Aujourd'hui, cette maladie est la cause de 2 à 3 millions de morts dans le monde et environ 6  
20 millions de nouveaux cas sont observés chaque année (Bouvet, 1994). Dans les pays développés *M. tuberculosis* est la cause la plus commune des infections mycobactériennes. En France il apparaît environ 10 000 nouveaux cas par an et parmi les maladies à déclaration  
25 obligatoire c'est la tuberculose qui comprend le plus grand nombre de cas. La vaccination par le BCG (Bacille de Calmette et Guérin), une souche avirulente dérivée de *M. bovis* et qui est très utilisée comme vaccin contre la tuberculose, est loin d'être efficace au sein de toutes les  
30 populations. Cette efficacité varie environ de 80 % dans les pays occidentaux comme l'Angleterre, à 0 % en Inde (résultats de l'enquête de la Commission de l'OMS sur la tuberculose).

La présente invention concerne des séquences de l'ADN et du génome de *M. tuberculosis* et de *M. leprae* et des produits dérivés de ces séquences.



une tuberculose, rendent nécessaire la mise au point de méthodes rapides, spécifiques et fiables pour le diagnostic de la tuberculose et la mise au point de nouveaux vaccins. Par exemple, une étude épidémiologique réalisée en Floride, et dont les résultats ont été publiés en 1993 dans AIDS 5 thérapies, a montré que 10 % des malades atteints de SIDA sont atteints de tuberculose au moment du diagnostic du SIDA ou 18 mois avant celui-ci. Chez ces malades, la tuberculose apparaît dans 60 % des cas sous une forme 10 disséminée donc non repérable par les critères de diagnostic classiques comme la radiographie pulmonaire ou l'analyse de crachats.

Actuellement, une certitude sur le diagnostic apporté 15 par la mise en évidence de bacilles cultivables dans un prélèvement provenant du malade n'est obtenue que pour moins de la moitié des cas de tuberculose, même dans les cas de tuberculose pulmonaire. Le diagnostic de la tuberculose et des autres mycobactéries apparentées est 20 donc difficile à réaliser, et cela pour différentes raisons : les mycobactéries sont souvent présentes en faible quantité, leur temps de génération est très long (24h pour *M. tuberculosis*) et leur culture est difficile. (Bates et al., 1986).

25 D'autres techniques sont utilisables en clinique, pour identifier une infection mycobactérienne :

a) L'identification directe des microorganismes au microscope ; cette technique est rapide, mais ne permet pas 30 l'identification de l'espèce mycobactérienne observée et manque de sensibilité (Bates, 1987).

Les cultures, lorsqu'elles sont positives, ont une

cultures répétées (Bates, 1979 ; Bates et al., 1986)) et coûteuse.

b) Les techniques sérologiques peuvent s'avérer utiles dans certaines conditions, mais leur utilisation est parfois limitée par leur sensibilité et/ou leur spécificité faibles (Daniel et al., 1987).

c) La présence de mycobactéries au sein d'un échantillon biologique peut aussi être déterminée par hybridation moléculaire avec de l'ADN ou de l'ARN en utilisant des sondes d'oligonucléotides spécifiques des séquences recherchées (Kiehn et al., 1987 ; Roberts et al., 1987 ; Drake et al., 1987). Plusieurs études ont montré l'intérêt de cette technique pour le diagnostic des infections à mycobactéries. Les sondes utilisées sont constituées d'ADN, d'ARN ribosomal ou de fragments d'ADN mycobactériens provenant de banque de gènes. Le principe de ces techniques repose sur le polymorphisme des séquences nucléotidiques des fragments utilisés ou sur le polymorphisme des régions avoisinantes. Dans tous les cas, elles nécessitent l'utilisation de cultures et ne sont pas applicables directement sur les échantillons biologiques.

La faible quantité de mycobactéries présentes au sein d'un échantillon biologique et en conséquence la quantité faible d'ADN cible à détecter dans cet échantillon peut nécessiter le recours à une amplification spécifique in vitro de l'ADN cible avant sa détection à l'aide de la sonde nucléotidique et en utilisant des techniques d'amplification in vitro telles que la PCR (amplification en chaîne à la polymérase) (Mullis et al., 1986) ou la LAMP (amplification isothermale à la polymérase) (Nottingham et al., 1990).

second temps à l'aide d'une sonde oligonucléotidique capable de s'hybrider spécifiquement à l'ADN amplifié.

Un test de détection de mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*, par hybridation sandwich (test utilisant une sonde de capture et une sonde de détection) a été décrit par Chevrier et al. en 1993. le complexe de *Mycobacterium tuberculosis* est un groupe de mycobactéries qui comprend *M. bovis*-BCG, *M. bovis*, *M. tuberculosis*, *M. africanum* et *M. microti*.

10 Un procédé de détection de faibles quantités de mycobactéries, appartenant au complexe tuberculosis, par amplification génique et hybridation directement sur des échantillons biologiques a été mis au point. Ledit procédé  
15 utilise la séquence d'insertion IS6110 (Brevet européen EP 0 490 951 B1). Thierry et al. ont décrit en 1990 une séquence spécifique du complexe *Mycobacterium tuberculosis* et nommée IS 6110. Certains auteurs ont proposé d'amplifier spécifiquement l'ADN provenant de  
20 *Mycobacterium* en utilisant des amorces nucléiques dans une méthode d'amplification, telle que la réaction de polymérase en chaîne (PCR). Patel et al. ont décrit en 1990 l'utilisation de plusieurs amorces nucléiques choisies à partir d'une séquence connue en tant que sonde dans  
25 l'identification de *M. tuberculosis*. Cependant, la longueur des fragments obtenus en utilisant ces amorces était différente de la longueur théorique attendue et plusieurs fragments de taille variable étaient obtenus. De plus, les auteurs ont observé l'absence d'hybridation des produits  
30 amplifiés avec le plasmide ayant servi à déterminer les amorces. Ces résultats indiquent que ces amorces

détection de *M. tuberculosis*, de grande sensibilité, par amplification d'un fragment d'ADN de *M. tuberculosis* localisé au sein de la séquence IS6110 (Brevet européen EP 461 045) à l'aide d'amorces générant des fragments d'ADN  
5 amplifié de longueur constante, même lorsque le choix des amorces conduisait à l'amplification de fragments longs (de l'ordre de 1000 à 1500 bases) où le risque d'interruption de la polymérisation est élevée en raison des effets de la structure secondaire de la séquence. D'autres amorces  
10 spécifiques de la séquence IS6110 sont décrites dans le brevet européen N° EP-0490 951.

Les inventeurs ont montré (résultats non publiés) que certains isolats cliniques de *Mycobacterium tuberculosis*  
15 étaient exempts de la séquence d'insertion IS6110 et ne pouvaient donc être détectés à l'aide des oligonucléotides spécifiques de cette séquence pouvant conduire ainsi à des résultats de diagnostic faussement négatifs. Ces résultats confirment une observation similaire faite par Yuen et al.  
20 en 1993. L'impossibilité de détecter ces souches pathogènes potentiellement présentes dans un échantillon biologique prélevé sur un patient est ainsi susceptible de conduire à des difficultés voire des erreurs de diagnostic. La disponibilité de plusieurs séquences spécifiques du Bacille de la tuberculose, à l'intérieur desquelles des amorces  
25 appropriées pour l'amplification seront choisis, est importante. La séquence LP424 décrite ici pourra être utilisée.

30 *M. bovis* et *M. tuberculosis*, les agents causaux de la tuberculose, sont des bactéries facultatives intracellulaires.

éradiquer l'invasion par des microorganismes. Ces agents sont capables de moduler l'évolution normale de leur phagosome et de les empêcher de se différencier en un compartiment acide riche en hydrolase (Clemens, 1979 ; Clemens et al., 1996; Sturgill-Koszycki et al., 1994 et Xu et al., 1994). Cependant, cette modulation n'est possible que si la bactérie est vivante au sein du phagosome, suggérant que des composés synthétisés de manière active et/ou sécrétés à l'intérieur de la cellule font partie de ce mécanisme. Des protéines exportées sont probablement impliquées dans ce mécanisme. En dépit des problèmes majeurs de santé liés à ces organismes pathogènes, on sait peu de choses sur leurs protéines exportées et/ou sécrétées. Des analyses en SDS-PAGE de filtrat de culture de *M. tuberculosis* montrent au moins 30 protéines sécrétées (Altschul et al., 1990 ; Nagai et al., 1991 et Young et al., 1992). Certaines d'entre elles ont été caractérisées, leurs gènes clonés et séquencés (Borremans et al., 1989 ; Wiker et al., 1992 et Yamaguchi et al., 1989). D'autres, bien qu'il s'agisse d'antigènes immunodominants d'importance majeure pour induire une immunité protectrice (Anderson et al., 1991 et Orme et al., 1993), ne sont pas totalement identifiés. En outre, il est probable que de nombreuses protéines exportées restent fixées sur la membrane cellulaire et par conséquent ne soient pas présentes dans les surnageants de culture. Il a été montré que les protéines localisées à la surface externe de diverses bactéries pathogènes, telles que l'invasine de 103 kDa de *Yersinia pseudotuberculosis* (Isberg et al., 1987) ou l'internaline de 86 kDa de *Listeria monocytogenes* (Gaillard et al., 1991 et Brameri et al., 1992) jouent un rôle important dans les interactions avec les cellules hôtes.

réponse protectrice contre cette infection. Ces protéines pourraient revêtir un intérêt certain pour la préparation de vaccins.

5           Récemment, il a été décrit l'adaptation aux mycobactéries d'une méthodologie génétique pour l'identification et la sélection phénotypique de protéines exportées (Lim et al., 1995). Cette méthode utilise la phosphatase alcaline (PhoA) périplasmique d'*E. coli*. Un  
10 vecteur plasmidique a été construit permettant la fusion de gènes entre un gène *PhoA* tronqué et des gènes codant pour des protéines exportées (Manoil et al., 1990).

Par cette méthode, il a pu être identifié un gène  
15 de *M. tuberculosis* (*erp* (Berthet et al., 1995)) présentant des homologies avec une protéine exportée de 28 kDa de *M. leprae*, qui est une cible fréquente des réponses humorales de la forme lépromateuse de la lèpre. Une protéine présentant des motifs aminocides caractéristiques de la  
20 désaturase de plante (*des*) a aussi été caractérisée par la technique de fusion avec *PhoA*.

Cependant, cette méthode génétique d'identification de protéines exportées ne permet pas d'évaluer facilement  
25 l'expression intracellulaire des gènes correspondants. Une telle évaluation est d'une importance primordiale à la fois pour la sélection de bons candidats vaccins et pour la compréhension des interactions entre les bactéries et leurs cellules hôtes. L'induction de l'expression de facteur de  
30 virulence par contact de cellule cible pathogène a été décrite. C'est le cas par exemple pour les facteurs de virulence de *M. tuberculosis*.

de surface.

Compte tenu de ce qui précède, il existe aujourd'hui un grand besoin de développer de nouveaux vaccins contre les mycobactéries pathogènes ainsi que de nouveaux tests de diagnostic spécifiques, fiables et rapides. Ces développements nécessitent la mise au point d'outils spécifiques encore plus performants permettant, d'une part, d'isoler ou d'obtenir des séquences de nouveaux polypeptides spécifiques, notamment immunogènes, et, d'autre part, de mieux comprendre le mécanisme des interactions entre les bactéries et leurs cellules hôtes comme notamment l'induction de l'expression de facteur de virulence. Ceci est précisément l'objet de la présente invention.

Les inventeurs ont défini et réalisé dans ce but de nouveaux vecteurs permettant le criblage, le clonage et/ou l'expression de séquences d'ADN de mycobactéries afin d'identifier parmi ces séquences, des acides nucléiques codant pour des protéines d'intérêt, de préférence des protéines exportées, pouvant être localisées sur la membrane bactérienne et/ou sécrétées, et d'identifier parmi ces séquences celles qui sont induites ou réprimées lors de l'infection (croissance intracellulaire).

#### Description

La présente invention décrit l'utilisation du gène rapporteur *phoA* chez les mycobactéries. Il permet d'identifier des systèmes d'expression et d'exportation dans un contexte mycobactérien.

début du gène, ses régions régulatrices et son régulateur seront clonés ce qui permettra d'observer une régulation. Si cette régulation est positive, le clonage du régulateur constituera un avantage pour observer l'expression et l'exportation.

Dans le contexte de l'invention, on entend par mycobactérie toutes les mycobactéries appartenant aux diverses espèces énumérées par Wayne L. G. and Kubica G. P. (1980). Family Mycobacteriaceae in Bergey's manual of systematic bacteriology, J. P. Butler Ed. (Baltimore USA : Williams et Wilkins P. 1436-1457).

Dans certains cas les gènes clonés sont soumis dans leur hôte d'origine à une régulation négative rendant l'observation de l'expression et de l'exportation difficile chez l'hôte d'origine. Dans ce cas, le clonage du gène en absence de son régulateur négatif, dans un hôte ne le contenant pas, constituera un avantage.

L'invention vise aussi de nouveaux polypeptides et de nouveaux polynucléotides de mycobactéries ayant pu être isolés au moyen des vecteurs précédents et susceptibles d'entrer dans la réalisation de compositions pour la détection d'une infection par des mycobactéries, ou pour la protection contre une infection due à des mycobactéries ou pour la recherche d'inhibiteurs comme cela est décrit précédemment pour DP428.

L'invention a donc pour objet un vecteur recombinant de criblage, de clonage et/ou d'expression, caractérisé en ce qu'il est :



3) une cassette rapporteur comprenant :

- a) un site de clonage multiple (polylinker),
- b) éventuellement un terminateur de transcription  
5 actif chez les mycobactéries, en amont du polylinker,
- c) une séquence nucléotidique codante issue d'un gène  
codant pour un marqueur d'expression, d'exportation et/ou  
de sécrétion de protéine, ladite séquence nucléotidique  
étant dépourvue de son codon d'initiation et de ses  
10 séquences de régulation, et
- d) une séquence nucléotidique codante issue d'un gène  
codant pour un marqueur d'activité de promoteurs contenus  
dans le même fragment, ladite séquence nucléotidique étant  
pourvue de son codon d'initiation. Eventuellement, le  
15 vecteur recombinant contient également un réplicon  
fonctionnel chez E. coli.

De manière préférée, le marqueur d'exportation et/ou  
de sécrétion est placé dans la même orientation que le  
20 marqueur d'activité de promoteurs.

Préférentiellement, le vecteur recombinant de criblage  
selon l'invention comprendra, en outre, un terminateur de  
transcription placé en aval du marqueur d'activité de  
25 promoteurs, ce qui est de nature à permettre l'obtention de  
transcrits courts qui se révèlent plus stables et qui, par  
conséquent, permettent un plus haut niveau d'expression des  
produits de traduction.

30 Le marqueur d'exportation et/ou de sécrétion est une  
séquence de nucléotides dont l'expression suivie de  
l'exportation et/ou de la sécrétion dépend des éléments de  
régulation de la cellule.

transcription, une séquence comprenant le site de liaison au ribosome (RBS), les séquences responsables de l'exportation et/ou la sécrétion telles que la séquence dite séquence signal.

5

Un premier marqueur intéressant d'exportation et/ou d'expression est une séquence codante issue du gène *phoA*. Le cas échéant, elle est tronquée de telle façon que l'activité phosphatase alcaline est cependant susceptible d'être restaurée lorsque la séquence codante tronquée est placée sous le contrôle d'un promoteur et d'éléments de régulation appropriés.

10 D'autres marqueurs d'exposition, d'exportation et/ou de sécrétion peuvent être utilisés. On citera à titre d'exemples une séquence du gène  $\beta$ -agarase, de la nucléase d'un staphylocoque ou d'une  $\beta$ -lactamase.

20 Parmi les marqueurs intéressants d'activité de promoteurs contenus dans le même fragment, on préfère une séquence codante issue du gène *lux* de luciférase de luciole pourvue de son codon d'initiation.

25 D'autres marqueurs d'activité de promoteurs contenus dans le même fragment peuvent être utilisés. On citera à titre d'exemples une séquence du gène de la GFP (Green Fluorescent Protein).

30 Le terminateur de transcription doit être fonctionnel chez les mycobactéries. Un terminateur avantageux est à cet égard le terminateur du phage T4 (TTA). D'autres terminateurs appropriés sont connus.

Un vecteur particulièrement préféré pour la réalisation de l'invention est un plasmide choisi parmi les plasmides suivants déposés à la CNCM (Collection Nationale de Cultures de Microorganismes, 25 rue de Docteur Roux, 5 75724 Paris cedex 15, France) :

a) pJVEDa déposé à la CNCM sous le N° I-1797, le 12/12 1996,

b) pJVEDb déposé à la CNCM sous le N° I-1906, le 25 10 juillet 1997,

c) pJVEDc déposé à la CNCM sous le N° I-1799, le 12/12 1996.

Pour la sélection, ou l'identification de séquences 15 d'acides nucléiques de mycobactéries codant pour des polypeptides susceptibles d'être incorporés dans des compositions immunogènes, ou antigéniques pour la détection d'une infection, ou susceptibles d'induire ou de réprimer un facteur de virulence de mycobactéries, le vecteur de 20 l'invention comprendra, en l'un des sites de clonage multiple du polylinker, une séquence de nucléotides d'une mycobactérie chez laquelle on détecte la présence de séquences correspondant à des polypeptides exportés et/ou 25 sécrétés pouvant être induits ou réprimés lors de l'infection, ou encore exprimés ou produits de façon constitutive, leurs séquences promotrices et/ou régulatrices associées susceptibles de permettre ou de favoriser l'exportation et/ou la sécrétion desdits polypeptides d'intérêt, ou tout ou partie de gènes 30 d'intérêt codant pour lesdits polypeptides.

De préférence, cette séquence est obtenue par fragmentation d'un gène.

Les vecteurs de l'invention peuvent en effet également être utilisés pour déterminer la présence de séquences d'intérêt, de préférence correspondant à des protéines exportées et/ou sécrétées, et/ou capables d'être induites ou réprimées ou produites de façon constitutive lors de l'infection, notamment lors de la phagocytose par les macrophages, et selon ce qui a été exposé précédemment, chez des mycobactéries telles que *M. africanum*, *M. bovis*, *M. avium* ou *M. leprae* dont on aura traité l'ADN ou l'ADNc par fragmentation physique ou avec des enzymes déterminées.

Selon un premier mode de réalisation de l'invention la digestion enzymatique de l'ADN génomique ou de l'ADN complémentaire est effectuée à partir de *M. tuberculosis*.

De préférence cet ADN est digéré avec une enzyme telle que *sau3A*, *BclI*, *BglII*.

D'autres enzymes de digestion telles que *ScaI*, *ApaI*, *SacII*, *KpnI* ou encore des nucléases ou des polymérases, peuvent naturellement être mises en oeuvre, dès lors qu'elles permettent l'obtention de fragments dont les extrémités peuvent être insérées dans l'un des sites de clonage du polylinker du vecteur de l'invention.

Le cas échéant, des digestions avec différentes enzymes seront effectuées simultanément.

Les vecteurs recombinants préférés pour la réalisation de l'invention sont choisis parmi les vecteurs recombinants suivants déposés à la CNOC :

a) p6P7 déposé le 29 janvier 1997 à la CNOC sous le N° 1814,

b) p8A1 déposé le 29 janvier 1997 à la CNOC sous le N° 1815.

- e) pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818,  
f) p5B5 déposé le 28 janvier 1997 à la CNCM sous le N°I-1819,  
5 g) p1C7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1820,  
h) p2D7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1821,  
10 i) p1B7 déposé le 31 janvier 1997 à la CNCM sous le N°I-1843,  
j) pJVED/M. tuberculosis déposé le 25 juillet 1997 à la CNCM sous le N° I-1907,  
k) pM1C25 déposé le 4 août 1998 à la CNCM sous le n°I-2062.

15

Parmi les plus préférés, on préfère le vecteur recombinant pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818, et le vecteur pM1C25 déposé le 4 août 1998 à la CNCM sous le N° I-2062.

20

L'invention a également pour objet un procédé de criblage de séquences de nucléotides issues de mycobactéries pour déterminer la présence de séquences correspondant à des polypeptides exportés et/ou sécrétés  
25 pouvant être induits ou réprimés lors de l'infection, leurs séquences promotrices et/ou régulatrices associées susceptibles notamment de permettre ou de favoriser l'exportation et/ou la sécrétion de certains polypeptides d'intérêt, ou tout ou partie de gènes d'intérêt codant pour  
30 lesdits polypeptides, caractérisé en ce qu'il met en œuvre un vecteur recombinant selon l'invention.

L'invention concerne aussi un procédé de criblage

Le présent document a été déposé en français et en anglais. Les versions anglaise et française ont été comparées et il a été constaté qu'elles étaient substantiellement concordantes. Les versions anglaise et française ont été comparées et il a été constaté qu'elles étaient substantiellement concordantes.

déterminée et la récupération des fragments obtenus ;

b) l'insertion des fragments obtenus à l'étape a) dans un site de clonage, compatible le cas échéant avec l'enzyme de l'étape a), du polylinker d'un vecteur selon l'invention ;

c) si besoin, l'amplification desdits fragments contenus dans le vecteur, par exemple par répllication de ce dernier après insertion du vecteur ainsi modifié dans une cellule déterminée, de préférence *E coli* ;

d) la transformation des cellules hôtes par le vecteur amplifié à l'étape c), ou en l'absence d'amplification, par le vecteur de l'étape b) ;

e) la culture des cellules hôtes transformées dans un milieu permettant la mise en évidence du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs contenu dans le vecteur ;

f) la détection des cellules hôtes positives (colonies positives) pour l'expression du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs ;

g) l'isolement de l'ADN des colonies positives et l'insertion de cet ADN dans une cellule identique à celle de l'étape c) ;

h) la sélection des insertions contenues dans le vecteur, permettant l'obtention de clones positifs pour le marqueur d'exportation et/ou de sécrétion, et /ou pour le marqueur d'activité de promoteurs ;

i) l'isolement et la caractérisation des fragments d'ADN de mycobactéries contenues dans ces insérats.

Dans l'un des modes de réalisation préférés du procédé de criblage selon l'invention, les cellules hôtes

marqueur d'activité de promoteurs lorsque lesdites cellules hôtes sont phagocytées par des cellules du type macrophagique.

- 5 De manière plus spécifique, on compare la stimulation de l'expression du marqueur d'activité de promoteurs chez des cellules hôtes placées en culture axénique (cellules hôtes seules en culture) à la stimulation de l'expression du marqueur d'activité de promoteurs chez des cellules  
10 hôtes cultivées en présence de macrophages et ainsi phagocytées par ces derniers.

La sélection de cellules hôtes positives pour le marqueur d'activité de promoteurs peut être réalisée dès  
15 l'étape c) du procédé de criblage décrit ci-dessus, ou encore après l'une quelconque des étapes f), g), h) ou i), c'est-à-dire une fois que les cellules hôtes ont été sélectionnées positivement pour le marqueur d'exportation et/ou de sélection.

20 La mise en oeuvre de ce procédé permet la construction de banques d'ADN comportant des séquences correspondant à des polypeptides susceptibles d'être exportés et/ou sécrétés, et/ou susceptibles d'être induits ou réprimés  
25 lors de l'infection lorsqu'ils sont produits au sein de mycobactéries recombinantes. L'étape i) du procédé peut comprendre une étape de séquençage des insertions sélectionnées.

30 De préférence, dans le procédé selon l'invention, le vecteur utilisé est choisi parmi les plasmides pJVEH4 (CNRM, N° 1-1727), pJVEDB (CNRM, N° 1-1906), pJVEDc (CNRM, N° 1-1799) ou pJVEH4 modifié.

le procédé de criblage est caractérisé en ce que les séquences de mycobactéries sont issues d'une mycobactérie pathogène, par exemple de *M. tuberculosis*, *M. bovis*, *M. avium*, *M. africanum* ou *M. leprae*.

5

L'invention comprend également une banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie, caractérisée en ce qu'elle est obtenue par un procédé comprenant les étapes a) et b), ou a), b) et c) du procédé précédant selon l'invention, de préférence une banque d'ADN  
10 génomique ou d'ADNc complémentaire d'ARNm de mycobactéries pathogènes, de préférence de mycobactéries appartenant au groupe du complexe *Mycobacterium tuberculosis*, de préférence de *Mycobacterium tuberculosis*.

15

Dans la présente invention, on entend désigner par "séquences nucléiques" ou "séquences d'acides aminés" SEQ ID N° X à SEQ ID N° Y, où X et Y peuvent représenter indépendamment un nombre ou un caractère alphanumérique,  
20 respectivement l'ensemble des séquences nucléiques ou l'ensemble des séquences d'acides aminés représentées par les figures X à Y, extrémités comprises.

Par exemple, les séquences nucléiques ou les séquences  
25 d'acides aminés SEQ ID N° 1 à SEQ ID N° 4N sont respectivement les séquences nucléiques ou les séquences d'acides aminés représentées par les figures 1 à 4N, c'est-à-dire respectivement les séquences nucléiques ou les séquences d'acides aminés SEQ ID N° 1, SEQ ID N° 1A', SEQ ID N° 1B', SEQ ID N° 1C', SEQ ID N° 1D, SEQ ID N° 1E, SEQ ID N° 2, SEQ ID N° 3A, SEQ ID N° 3B, SEQ ID N° 3C, SEQ ID N° 4A, SEQ ID N° 4B, SEQ ID N° 4C, SEQ ID N° 4A', SEQ ID N° 4B', SEQ ID N° 4C', SEQ ID N° 4D, SEQ ID N° 4E, SEQ ID N° 4F, SEQ ID N° 4G, SEQ ID N° 4H, SEQ ID N° 4I, SEQ ID N° 4J, SEQ ID N° 4K, SEQ ID N° 4L, SEQ ID N° 4M, SEQ ID N° 4N.



nucléotidiques de mycobactéries ou comprenant des séquences nucléotidiques de mycobactéries sélectionnées après la réalisation du procédé selon l'invention ci-dessus décrit.

5

De préférence, ladite mycobactérie est choisie parmi *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. avium*, *M. leprae*, *M. paratuberculosis*, *M. kansasii* ou *M. xenopi*.

10

On préfère les séquences nucléotidiques de mycobactéries ou comprenant une séquence nucléotidique de mycobactérie, ladite séquence nucléotidique de mycobactérie étant choisie parmi les séquences de fragments d'ADN de mycobactérie de séquence nucléique SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 27C, SEQ ID N° 29 et SEQ ID N° 31A à SEQ ID N° 50F, respectivement représentées par les figures 1 à 24C (planches 1 à 150), par les figures 27A à 27C (planches 152 à 154), par la figure 29 (planche 156) et par les figures 31A à 50F (planches 158 à 275).

20

Selon un mode de réalisation particulier de l'invention, des séquences préférées sont par exemple les fragments d'ADN de mycobactéries de séquence SEQ ID N°1, SEQ ID N°3A, SEQ ID N°5A, SEQ ID N°6A, SEQ ID N°7A, SEQ ID N°8A, SEQ ID N°9A, SEQ ID N°10A, SEQ ID N°27A et SEQ ID N°29 contenus respectivement dans les vecteurs pDP478 (CNCM, N°1-1818), p6D7 (CNCM, N°1-1814), p5F6 (CNCM, N°1-1816), p2A29 (CNCM, N°1-1817), p5B5 (CNCM, N°1-1815), p1C7 (CNCM, N°1-1820), p2D7 (CNCM, N°1-1821), p1B7 (CNCM, N°1-1843), p5A3 (CNCM, N°1-1815) et pM1C25 (CNCM, N°1-2620).

30

L'invention concerne également un acide nucléique comprenant la totalité d'une des séquences ci-dessus.

Il est bien entendu que les séquences ci-dessus peuvent être

isolé par exemple de la façon suivante :

- a) préparation d'une banque de cosmides à partir de l'ADN de *M. tuberculosis*, par exemple selon la technique décrite par Jacobs et al., 1991;
- b) hybridation de tout ou partie d'un acide nucléique sonde de séquence choisie par exemple parmi SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 27C, SEQ ID N° 29 et SEQ ID N° 31A à SEQ ID N° 50F avec les cosmides de la banque préalablement préparée à l'étape a) ;
- c) sélection des cosmides hybridant avec l'acide nucléique sonde de l'étape b) ;
- d) séquençage des inserts d'ADN des clones sélectionnés à l'étape c) et identification du cadre de lecture ouvert complet ;
- e) le cas échéant, clonage des inserts séquencés à l'étape d) dans un vecteur d'expression et/ou de clonage approprié.

Les acides nucléiques comprenant la totalité du cadre de lecture ouvert des séquences SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 27C, SEQ ID N° 29 et SEQ ID N° 31A à SEQ ID N° 50F sont parmi les acides nucléiques préférés.

La présente invention permet de déterminer un fragment de gène codant pour un polypeptide exporté. La comparaison avec la séquence du génome publiée par Cole et al. (Cole et al., 1998, Nature, 393, 537-544) permet de déterminer le gène en entier portant la séquence identifiée selon la présente invention.

selon l'invention et s'étendant d'une part en 5' de ces séquences jusqu'au premier codon d'initiation de la traduction (ATG ou GTG) ou même jusqu'au premier codon stop, et d'autre part en 3' de ces séquences jusqu'au codon stop suivant, et ceci dans l'une quelconque des trois phases de lecture possibles.

Les séquences nucléotidiques complémentaires des séquences ci-dessus selon l'invention font également partie de l'invention.

Par polynucléotide de séquence complémentaire d'une séquence nucléotidique selon l'invention, on entend toute séquence d'ADN ou d'ARN dont les nucléotides sont complémentaires de ceux de ladite séquence selon l'invention et dont l'orientation est inversée.

Les fragments nucléotidiques des séquences ci-dessus selon l'invention notamment utiles en tant que sondes ou amorces font également partie de l'invention.

L'invention concerne aussi les polynucléotides caractérisés en ce qu'ils comprennent un polynucléotide choisi parmi :

- a) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide selon l'invention,
- b) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide selon l'invention,
- c) un polynucléotide hybridant dans des conditions de forte stringence avec une séquence de polynucléotide selon l'invention,
- d) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide selon l'invention.

Lorsque la séquence codante issue du gène marqueur d'exportation et/ou de sécrétion est une séquence issue du gène *phoA*, l'exportation et/ou la sécrétion du produit du gène *phoA*, le cas échéant tronqué, n'est obtenue que lorsque cette séquence est insérée en phase avec la séquence ou élément de régulation de l'expression de la production de polynucléotides et sa localisation placée en amont, qui contient les éléments contrôlant l'expression, l'exportation et/ou la sécrétion issus de séquence de mycobactéries.

Les vecteurs recombinants de l'invention peuvent bien entendu comprendre des sites de clonage multiples décalés de un ou deux nucléotides par rapport à un vecteur selon l'invention, permettant ainsi d'exprimer le polypeptide correspondant au fragment d'ADN de mycobactérie inséré et susceptible d'être traduit selon l'un des trois cadres de lecture possibles.

20

Par exemple les vecteurs préférés pJVEDb et pJVEDc de l'invention se distinguent du vecteur préféré pJVEDa par un décalage respectif de un et de deux nucléotides au niveau du site de clonage multiple.

25

Ainsi, les vecteurs de l'invention sont capables d'exprimer chacun des polypeptides susceptibles d'être codés par un fragment d'ADN de mycobactérie inséré. Cesdits polypeptides, caractérisés en ce qu'ils sont donc susceptibles d'être exportés et/ou sécrétés, et/ou induits ou réprimés, ou exprimés de façon constitutive lors de

30

ID N° 27A à SEQ ID N° 28 et SEQ ID N° 30 à SEQ ID N° 50F et représentées respectivement par les figures 1 à 24C (planches 1 à 150), les figures 27A à 28 (planches 152 à 155) et les figures 30 à 50F (planches 157 à 275).

5

Font également partie de l'invention, les fragments ou fragments biologiquement actifs ainsi que les polypeptides homologues desdits polypeptides. Fragment, fragment biologiquement actif et polypeptides homologue de polypeptide, étant tels que définis ci après dans la description.

L'invention concerne également les polypeptides comprenant un polypeptide ou un de leurs fragments selon l'invention.

15

..

L'invention a aussi pour objet des mycobactéries recombinantes contenant un vecteur recombinant selon l'invention décrit précédemment. Une mycobactérie préférée est une mycobactérie du type *M. smegmatis*.

20

*M. smegmatis* permet avantageusement de tester l'efficacité de séquences de mycobactéries, pour le contrôle de l'expression, de l'exportation et/ou de la sécrétion, et/ou de l'activité de promoteurs d'une séquence donnée, par exemple d'une séquence codant pour un marqueur tel que la phosphatase alcaline et/ou la luciférase.

25

Une autre mycobactérie préférée est une mycobactérie du type *M. bovis*, par exemple la souche BCG utilisée actuellement pour la vaccination contre la tuberculose.

30

Une autre mycobactérie préférée est une souche de *M. tuberculosis*, *M. bovis* ou *M. goodii*.

présente chez toutes les souches testées de mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*. Ce polynucléotide, dénommé DP428 contient un cadre ouvert de lecture (ORF) codant pour un polypeptide d'environ 12 kD.

5 Le cadre de lecture ouvert (ORF) codant pour le polypeptide DP428 s'étend du nucléotide en position nt 941 au nucléotide en position nt 1351 de la séquence SEQ ID N° 2, le polypeptide DP428 ayant la séquence en acides aminés SEQ ID N° 28 suivante :

10 MKTGTATRRRLAVLIALLALPGAVALIAEPSATGASDPCAASEVARTVGSVAKSMGD  
YLDSPETNQVMTAVLQQQVGPGSVASLKAHFEANPKVASDLHALSQPI.TDLSTRCSLP  
ISGLQAIGLMQAVQGARR.

Ce poids moléculaire (PM) correspond au PM théorique de la protéine mature obtenue après clivage de la séquence  
15 signale, le PM de la protéine ou polypeptide DP428 étant d'environ 10 kD après ancrage potentiel au peptidoglycane et coupure potentielle entre S et G du motif LPISG.

20 Ce polynucléotide inclut, d'une part, un cadre ouvert de lecture correspondant à un gène de structure et, d'autre part, les signaux de régulation de l'expression de la séquence codante en amont et en aval de cette dernière. Le polypeptide DP428 est composé d'un peptide signal, d'une  
25 région centrale hydrophile et d'une région C-terminale hydrophobe. Cette dernière se termine par deux résidus arginines (R), signal de rétention, et est précédé par un motif LPISG qui rappelle le motif LPXTG d'ancrage au peptidoglycane (Schneewind et al., 1995).

30 Par gène de structure aux fins de la présente invention, on entend un polynucléotide codant pour une protéine ou un peptide.

Le présent document a pour objet de décrire un gène de structure et un gène de régulation de la transcription.

Ainsi, l'invention concerne en particulier un polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques SEQ ID N°1 à SEQ ID N°2.

5

Plus particulièrement, l'invention concerne un polynucléotide caractérisé en ce qu'il comprend un polynucléotide choisi parmi .

- 10 a) un polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques SEQ ID N°1 à SEQ ID N°2,
- b) un polynucléotide dont la séquence nucléique est la séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses,
- 15 de la séquence SEQ ID N°1,
- c) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide défini en a) ou b),
- d) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide défini en a), b) ou c),
- 20 e) un polynucléotide hybridant dans des conditions de forte stringence avec une séquence de polynucléotide défini en a), b), c) ou d),
- f) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide défini en a), b), c), d) ou e).

25

On entend par séquence nucléotidique, polynucléotide ou acide nucléique, selon la présente invention, aussi bien un ADN double brin, un ADN simple brin que des produits de transcription desdits ADN.

30

Par pourcentage d'identité au sens de la présente invention, on entend un pourcentage d'identité entre les bases de deux polynucléotides.

Une hybridation dans des conditions de forte stringence signifie que les conditions de température et de force ionique sont choisies de telle manière qu'elles permettent le maintien de l'hybridation entre deux  
5 fragments d'ADN complémentaires.

A titre illustratif, des conditions de forte stringence de l'étape d'hybridation aux fins de définir les fragments polynucléotidiques décrits ci-dessus, sont  
10 avantageusement les suivantes :

l'hybridation est réalisée à une température préférentielle de 65°C, en présence de tampon commercialisé sous le nom de rapid-hyb buffer par Amersham (RPN 1636) et 100 µg/ml d'ADN de E.coli.

15 Les étapes de lavage peuvent, par exemple, être les suivantes :

- deux lavages de 10 min, préférentiellement à 65°C, dans un tampon 2 x SSC et 0,1% SDS;
- deux lavages de 10 min, préférentiellement à 65°C, dans  
20 un tampon 1 x SSC et 0,1% SDS;
- un lavage de 10 min, préférentiellement à 65°C, dans un tampon de 0,1 x SSC et 0,1% SDS.

1 x SSC correspond à 0,15 M NaCl et 0,05M citrate de Na et une solution de 1 x Denhardt correspond à 0,02% Ficoll, 0,02% de polyvinylpyrrolidone et 0,02% de sérum  
25 albumine bovine.

Avantageusement, un fragment nucléotidique répondant à la définition précédente aura au moins 8 nucléotides, de préférence au moins 12 nucléotides, et encore plus  
30 préférentiellement au moins 20 nucléotides consécutifs de la séquence dont il est issu. Les conditions d'hybridation de forte stringence



Pour les conditions de mise en oeuvre des enzymes de restriction dans le but d'obtenir des fragments nucléotidiques des polynucléotides selon l'invention, on se réfèrera avantageusement à l'ouvrage de Sambrook et al., 1989.

Avantageusement, un polynucléotide de l'invention contiendra au moins une séquence comprenant l'enchaînement de nucléotides allant du nucléotide en position nt 964 au nucléotide nt 1234 du polynucléotide de séquence SEQ ID N° 1.

La présente invention a pour objet un polynucléotide selon l'invention, caractérisé en ce que sa séquence nucléique hybride avec l'ADN de séquence de mycobactéries et préférentiellement avec l'ADN de séquence de mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*.

Le polynucléotide est défini par une séquence polynucléotidique telle que définie supra.

La présente invention a également pour objet un polypeptide issu d'une mycobactérie, caractérisé en ce qu'il est présent uniquement chez les mycobactéries appartenant au complexe de *Mycobacterium tuberculosis*.

L'invention concerne également un polypeptide caractérisé en ce qu'il comprend un polypeptide choisi parmi :

a) un polypeptide issu d'une mycobactérie appartenant au complexe de *Mycobacterium tuberculosis*.

b) un polypeptide issu d'une mycobactérie appartenant au complexe de *Mycobacterium tuberculosis*.

- c) un fragment d'au moins 5 acides aminés d'un polypeptide défini en a) ou b),
- d) un fragment biologiquement actif d'un polypeptide défini en a), b), ou c).

5

La présente invention a aussi pour objet un polypeptide dont la séquence d'acides aminés est comprise dans la séquence d'acides aminés SEQ ID N°1 ou SEQ ID N°2, ou un polypeptide de séquence d'acides aminés SEQ ID N°28.

10

Par polypeptide homologue, on entendra désigner les polypeptides présentant, par rapport au polypeptide naturel selon l'invention tel que le polypeptide DP428, certaines modifications comme en particulier une délétion, addition ou substitution d'au moins un acide aminé, une troncation, un allongement, une fusion chimérique, et/ou une mutation. Parmi les polypeptides homologues, on préfère ceux dont la séquence d'acides aminés présente au moins 30%, de préférence 50%, d'homologie avec les séquences d'acides aminés des polypeptides selon l'invention. Dans le cas d'une substitution, un ou plusieurs acides aminés consécutifs ou non consécutifs, sont remplacés par des acides aminés « équivalents ». L'expression acide aminé « équivalent » vise ici à désigner tout acide aminé susceptible d'être substitué à l'un des acides aminés de la structure de base sans cependant modifier essentiellement les propriétés immunogènes des peptides correspondants. En d'autres termes, les acides aminés équivalents seront ceux qui permettent l'obtention d'un polypeptide de séquence modifiée qui permet l'induction *in vivo* d'anticorps ou de cellules capables de reconnaître le polypeptide dont la

25  
30

(polypeptide DP428) ou l'un de ses fragments ci-dessus définis.

Ces aminoacyles équivalents peuvent être déterminés soit en s'appuyant sur leur homologie de structure avec les aminoacyles auxquels ils se substituent, soit sur les résultats des essais d'immunogénicité croisée auxquels les différents peptides sont susceptibles de donner lieu.

A titre d'exemple, on mentionnera les possibilités de substitutions susceptibles d'être effectuées sans qu'il en résulte une modification approfondie de l'immunogénicité des peptides modifiés correspondants, les remplacements, par exemple, de la leucine par la valine ou l'isoleucine, de l'acide aspartique par l'acide glutamique, de la glutamine par l'asparagine, de l'arginine par la lysine etc., les substitutions inverses étant naturellement envisageables dans les mêmes conditions.

Par fragment biologiquement actif, on entendra désigner en particulier un fragment de séquence d'acides aminés de polypeptide présentant au moins une des caractéristiques des polypeptides selon l'invention, notamment en ce qu'il est :

- capable d'être exporté et/ou sécrété par une mycobactérie, et/ou d'être induit ou réprimé lors de l'infection par la mycobactérie ; et/ou
- capable d'induire, de réprimer ou de moduler, directement ou indirectement, un facteur de virulence de mycobactérie ; et/ou
- capable d'induire une réaction d'immunogénicité dirigée contre les mycobactéries ; et/ou

capable d'être reconnu par un anticorps spécifique de mycobactérie.

Un polypeptide de l'invention, ou un de ses fragments, tels que définis précédemment, est susceptible d'être reconnu spécifiquement par les anticorps présents dans le sérum de patients infectés par des mycobactéries et préférentiellement des bactéries appartenant au complexe de *Mycobacterium tuberculosis* ou par des cellules de l'hôte infecté.

Font ainsi partie de l'invention les fragments du polypeptide dont la séquence d'acides aminés est comprise dans la séquence d'acides aminés de polypeptide selon l'invention, telle que les séquences d'acides aminés SEQ ID N°1 à SEQ ID N°2, ou un polypeptide de séquence d'acides aminés SEQ ID N°28, qui peuvent être obtenus par clivage dudit polypeptide par une enzyme protéolytique, telle que la trypsine ou la chymotrypsine ou la collagénase, ou par un réactif chimique, tel que le bromure de cyanogène (CNBr) ou encore en plaçant un polypeptide selon l'invention tel que le polypeptide DP428 dans un environnement très acide, par exemple à pH 2,5. Des fragments peptidiques préférés selon l'invention, pour une utilisation en diagnostic ou en vaccination, sont les fragments contenus dans des régions de polypeptide selon l'invention tel que le polypeptide DP428 susceptibles d'être naturellement exposées au solvant et de présenter ainsi des propriétés d'immunogénicité importante. De tels fragments peptidiques peuvent être préparés indifféremment par synthèse chimique, à partir d'hôtes transformés par un vecteur d'expression selon l'invention contenant un acide nucléique permettant l'expression desdits fragments, placé sous le contrôle des éléments de régulation et/ou d'expression appropriés, ou encore par clivage chimique ou enzymatique.

de la SEQ ID N°28. Les résultats de cette analyse sont présentés à la figure 54, où sont détaillés, pour chacun des acides aminés (AA) de position définie dans la SEQ ID N°28, l'indice d'hydrophilicité. Plus l'indice d'hydrophilicité est élevé, plus l'acide aminé considéré est susceptible d'être exposé au solvant dans la molécule native, et est en conséquence susceptible de présenter un degré d'antigénicité élevé. Ainsi, un enchaînement d'au moins sept acides aminés possédant un indice élevé d'hydrophilicité ( $>0,3$ ) peut constituer la base de la structure d'un peptide candidat immunogène selon la présente invention.

Les réponses immunitaires cellulaires de l'hôte à un polypeptide selon l'invention, peuvent être mises en évidence selon les techniques décrites par Colignon et al., 1996.

D'après les données de la carte d'hydrophilicité présentée à la Figure 54, les inventeurs ont pu définir des régions du polypeptide DP428 préférentiellement exposées au solvant, plus particulièrement la région localisée entre les acides aminés 55 et 72 de la séquence SEQ ID N° 28 et la région localisée entre les acides aminés 99 et 107 de la SEQ ID N° 28.

Les régions peptidiques du polypeptide DP428 définies ci-dessus peuvent être avantageusement mises en œuvre pour la réalisation des compositions immunogènes ou des compositions vaccinales selon l'invention.

Les polyméroléptides caractérisés en ce qu'ils sont

en ce que lesdites séquences sont choisies parmi les séquences d'acide nucléique de polynucléotides selon l'invention.

5 L'invention concerne en outre l'utilisation d'une séquence d'acide nucléique de polynucléotides selon l'invention comme sonde ou amorce, pour la détection et/ou l'amplification de séquence d'acide nucléique. Parmi ces séquences d'acide nucléique selon l'invention utilisables  
10 comme sonde ou amorce, on préfère les séquences d'acide nucléique de l'invention, caractérisée en ce que lesdites séquences sont des séquences, ou leur séquence complémentaire, comprises entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités  
15 incluses, de la séquence SEQ ID N°1.

Parmi les polynucléotides selon l'invention, utilisables comme amorces nucléotidiques, on préfère particulièrement les polynucleotides de séquence SEQ ID  
20 N°25 et SEQ ID N°26.

Les polynucléotides selon l'invention peuvent ainsi être utilisés pour sélectionner des amorces nucléotidiques, notamment pour la technique PCR (Erlich, 1989 ; Innis et  
25 al., 1990, et, Belis et al., 1991).

Cette technique nécessite la choix de paires d'oligonucléotides encadrant le fragment qui doit être amplifié. On peut, par exemple, se référer à la technique  
30 décrite dans le brevet américain U.S. N° 4 683 202. Ces amorces oligodésoxyribonucléotidiques ou leurs homologues nucléotidiques ont avantageusement une longueur d'au moins 8 nucléotides.

La présente invention est illustrée par les exemples suivants.

brin (+) [amorce aller] de la matrice et l'autre amorce est complémentaire du brin (-) [amorce retour]. Il est important que les amorces ne possèdent pas de structure secondaire ou de séquence complémentaire l'une de l'autre.

5 D'autre part, la longueur et la séquence de chaque amorce doivent être choisies de manière à ce que les amorces ne s'hybrident pas avec d'autres acides nucléiques provenant de cellules procaryotes ou eucaryotes, en particulier avec les acides nucléiques provenant d'autres mycobactéries

10 pathogènes, ni avec l'ADN ou l'ARN humain pouvant éventuellement contaminer l'échantillon biologique.

Les résultats présentés à la figure 51, montrent que la séquence codant pour le polypeptide DP428 (SEQ ID N° 28)

15 n'est pas retrouvée dans les ADNs de *M. fortuitum*, *M. simiae*, *M. avium*, *M. chelonae*, *M. flavescens*, *M. gordonae*, *M. marinum* et *M. kansasii*

Les fragments amplifiés peuvent être identifiés après

20 une électrophorèse en gel d'agarose ou de polyacrylamide, ou après une électrophorèse capillaire, ou encore après une technique chromatographique (filtration sur gel, chromatographie hydrophobe ou chromatographie échangeuse d'ions). La spécificité de l'amplification peut être

25 contrôlée par hybridation moléculaire en utilisant comme sondes les séquences nucléotidiques de polynucléotides de l'invention, des plasmides contenant ces séquences ou leurs produits d'amplification.

30 Les fragments nucléotidiques amplifiés peuvent être utilisés comme réactifs dans des réactions d'hybridation afin de mettre en évidence la présence, dans un échantillon

Parmi les polynucléotides selon l'invention, utilisables comme sondes nucléotidiques, on préfère tout particulièrement le fragment polynucléotidique comprenant la séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses, de la séquence de SEQ ID N°1.

Ces sondes et amplicons peuvent être marqués ou non par des éléments radioactifs ou par des molécules non radioactives, telles que des enzymes ou des éléments fluorescents..

L'invention vise également les fragments nucléotidiques susceptibles d'être obtenus par amplification à l'aide d'amorces selon l'invention.

D'autres techniques d'amplification de l'acide nucléique cible peuvent être avantageusement employées comme alternatives à la PCR.

La technique SDA (Strand Displacement Amplification) ou technique d'amplification à déplacement de brin (Walker et al., 1992) est une technique d'amplification isotherme dont le principe est fondé sur la capacité d'une enzyme de restriction de couper l'un des deux brins de son site de reconnaissance qui se trouve sous une forme hemiphosphorothioate et sur la propriété d'une ADN polymérase d'initier la synthèse d'un nouveau brin d'ADN à partir de l'extrémité 3'OH créée par l'enzyme de restriction et de déplacer le brin préalablement synthétisé qui se trouve en aval.

Les polynucléotides de l'invention, en particulier les amorces selon l'invention, peuvent également être mis en oeuvre dans d'autres procédés.



- la technique 3SR (Self-Sustained Sequence Replication), décrite par Guatelli et al. en 1990;
- la technique NASBA (Nucleic Acid Sequence Based Amplification), décrite par Kievitis et al. en 1991;
- 5 - la technique TMA (Transcription Mediated Amplification).

Les polynucléotides de l'invention peuvent aussi être employés dans des techniques d'amplification ou de modification de l'acide nucléique servant de sonde, telles que:

- 10 - la technique LCR (Ligase Chain Reaction), décrite par Landegren et al. en 1988 et perfectionnée par Barany et al. en 1991, qui emploie une ligase thermostable;
- la technique de RCR (Repair Chain Reaction), décrite par Segev en 1992;
- 15 - la technique CPR (Cycling Probe Reaction), décrite par Duck et al. en 1990;
- la technique d'amplification à la Q-beta-réplicase, décrite par Miele et al. en 1983 et perfectionnée notamment par Chu et al. en 1986, Lizardi et al. en 1988, puis par
- 20 Burg et al. ainsi que par Stone et al. en 1996.

Dans le cas où le polynucléotide cible à détecter est un ARN, par exemple un ARNm, on utilisera avantageusement, préalablement à la mise en oeuvre d'une réaction

25 d'amplification à l'aide des amorces selon l'invention ou à la mise en oeuvre d'un procédé de détection à l'aide des sondes de l'invention, une enzyme de type transcriptase inverse afin d'obtenir un ADN à partir de l'ARN contenu dans l'échantillon biologique. L'ADN obtenu servira alors

30 de cible pour les amorces ou les sondes mises en oeuvre dans le procédé d'amplification ou de détection selon l'invention.

moins 15 nucléotides, et de préférence au moins de 200 nucléotides.

Les sondes nucléotidiques selon l'invention sont capables de détecter des mycobactéries et  
5 préférentiellement des bactéries appartenant au complexe de *Mycobacterium tuberculosis*, plus précisément du fait que ces mycobactéries possèdent dans leur génome au moins une copie de polynucléotides selon l'invention. Ces sondes selon l'invention, sont capables, par exemple, de  
10 s'hybrider avec la séquence nucléotidique d'un polypeptide selon l'invention, plus particulièrement tout oligonucléotide hybridant avec la séquence SEQ ID N°1, codant pour le polypeptide DP428 de *M. tuberculosis*, et ne présentant pas de réaction d'hybridation croisée ou  
15 d'amplification (PCR) avec par exemple des séquences présentes chez des mycobactéries n'appartenant pas au complexe de *Mycobacterium tuberculosis*. Les sondes nucléotidiques selon l'invention hybrident spécifiquement avec une molécule d'ADN ou d'ARN de polynucléotide selon  
20 l'invention, dans des conditions d'hybridation de forte stringence telles que données sous forme d'exemple précédemment.

Les séquences non marquées peuvent être utilisées  
25 directement comme sondes, cependant les séquences sont généralement marquées par un élément radioactif (<sup>32</sup>P, <sup>33</sup>P, <sup>35</sup>S, <sup>125</sup>I), ou par une molécule non-radioactive (biotine, acétylaminofluorène, digoxigénine, 5-étromo-désoxyuridine, fluorescéine) pour obtenir des sondes utilisables pour de  
30 nombreuses applications.

Les exemples de marquages non radioactifs de l'anneau

Dans ce dernier cas, on pourra aussi utiliser l'une des méthodes de marquage décrites dans les brevets FR 2 422 956 et FR 2 518 755. La technique d'hybridation peut être réalisée de manières diverses (Matthews et al., 1988). La méthode la plus générale consiste à immobiliser l'acide nucléique extrait des cellules de mycobactéries sur un support (tel que nitrocellulose, nylon, polystyrène) et à incuber, dans des conditions bien définies, l'acide nucléique cible immobilisé avec la sonde. Après l'hybridation, l'excès de sonde est éliminé et les molécules hybrides formées sont détectées par la méthode appropriée (mesure de la radioactivité, de la fluorescence ou de l'activité enzymatique liée à la sonde).

Avantageusement, les sondes nucléotidiques marquées selon l'invention peuvent avoir une structure telle qu'elles rendent possible une amplification du signal radioactif ou non-radioactif. Un système d'amplification répondant à la définition ci-dessus comprendra des sondes de détection sous la forme d'un ADN ramifié, branché («branched DNA») telles que celles décrites par Urdea et al. en 1991. Selon cette technique, on utilisera avantageusement plusieurs types de sondes notamment une sonde de capture, afin d'immobiliser l'ADN ou l'ARN cible sur un support, et une sonde de détection. La sonde de détection lie un ADN «branché» présentant une structure ramifiée. L'ADN branché, à son tour, est capable de fixer des sondes oligonucléotidiques qui sont elles-mêmes couplées à des molécules de phosphatase alcaline. Puis l'activité de cette enzyme est mise en évidence grâce à un substrat chimio-luminescent, par exemple un dérivé du dioxétane phosphate.

sur un support et sert à capturer par hybridation spécifique l'acide nucléique cible obtenu à partir de l'échantillon biologique à tester. Si nécessaire, le support solide est séparé de l'échantillon et le duplex formé entre la sonde de capture et l'acide nucléique cible est ensuite détecté grâce à une seconde sonde, dite «sonde de détection», marquée par un élément facilement détectable.

Les fragments oligonucléotidiques peuvent être obtenus à partir des séquences selon l'invention, par coupure avec des enzymes de restriction, ou par synthèse chimique selon les méthodes classiques, par exemple selon la méthode décrite dans le brevet européen N° EP-0305929 (Millipore Corporation) ou encore par d'autres procédés.

Un mode de préparation approprié des acides nucléiques de l'invention comportant au maximum 200 nucléotides (ou 200 pb s'il s'agit d'acides nucléiques bicaténaires) comprend les étapes suivantes :

- la synthèse d'ADN en utilisant la méthode automatisée des bêta-cyanéthylphosphoramidite décrite en 1986,
- le clonage des acides nucléiques ainsi obtenus dans un vecteur approprié et la récupération des acides nucléiques par hybridation avec une sonde appropriée.

Un mode de préparation, par voie chimique, d'acides nucléiques selon l'invention de longueur supérieure à 200 nucléotides (ou 200 pb lorsqu'il s'agit d'acides nucléiques bicaténaires) comprend les étapes suivantes :

- l'assemblage d'oligonucléotides synthétisés chimiquement, pourvus à leur extrémité de sites de restrictions différents, dont les séquences sont compatibles avec l'enchaînement en acides aminés du polypeptide naturel selon le principe décrit en 1983,

Les sondes nucléotidiques utilisées pour la récupération des acides nucléiques recherchés dans les procédés sus-mentionnés, sont constituées généralement de 8 à 200 nucléotides de la séquence de polypeptide selon l'invention et sont susceptibles de s'hybrider avec l'acide nucléique recherché dans les conditions d'hybridation définies précédemment. La synthèse de ces sondes peut être effectuée selon la méthode automatisée des bêta cyanethylphosphoramidites décrite en 1986.

10

Les sondes oligonucléotidiques selon l'invention peuvent être mises en oeuvre au sein d'un dispositif de détection comprenant une banque matricielle d'oligonucléotides. Un exemple de réalisation d'une telle banque matricielle peut consister en une matrice d'oligonucléotides sondes fixés sur un support, la séquence de chaque sonde d'une longueur donnée étant située en décalage d'une ou plusieurs bases par rapport à la sonde précédente, chacune des sondes de l'arrangement matriciel étant ainsi complémentaire d'une séquence distincte de l'ADN ou l'ARN cible à détecter et chaque sonde de séquence connue étant fixée en une position prédéterminée du support. La séquence cible à détecter peut être avantageusement marquée radioactivement ou non radioactivement. Lorsque la séquence cible marquée est mise en contact avec le dispositif matriciel, celle-ci forme des hybrides avec les sondes de séquences complémentaires. Un traitement à la nucléase, suivi d'un lavage, permet d'éliminer les hybrides sonde-séquence cible qui ne sont pas parfaitement complémentaires. Du fait de la connaissance précise de la séquence d'une sonde à une position déterminée de la matrice, il est alors possible de

20  
25  
30

Une alternative à l'utilisation d'une séquence cible marquée peut consister en l'utilisation d'un support permettant une détection « bioélectronique » de l'hybridation de la séquence cible sur les sondes du support matrice, lorsque que ledit support est constitué ou comprend un matériau capable d'agir, par exemple, en tant que donneur d'électrons aux positions de la matrice auxquelles un hybride a été formé. Un tel matériau donneur d'électron est par exemple de l'or. La détection de la séquence nucléotidique de l'ADN ou ARN cible est alors déterminée par un dispositif électronique.

Un exemple de réalisation d'un biocapteur, tel que défini ci-dessus, est décrit dans la demande de brevet européen N° EP-0721 016 au nom de Affymax technologies N.V. ou encore dans le brevet américain N° US 5.202.231 au nom de Drmanac.

L'invention a aussi pour objet les polynucléotides hybrides résultant :

- soit de la formation d'une molécule hybride entre un ARN ou un ADN (ADN génomique ou ADNc) provenant d'un échantillon biologique avec une sonde ou une amorce selon l'invention.
- soit de la formation d'une molécule hybride entre un ARN ou un ADN (ADN génomique ou ADNc) provenant d'un échantillon biologique avec un fragment nucléotidique amplifié à l'aide d'un couple d'amorces selon l'invention.

Par ADNc au sens de la présente invention, on entend une molécule d'ADN obtenue en faisant agir une enzyme de type transcriptase inverse sur une molécule d'ARN, en particulier une molécule d'ARN messager (ARNm), selon les techniques décrites dans l'art de l'état de la technique.

La présente invention a pour objet :

un procédé de détection d'une séquence cible par l'intermédiaire d'un support matrice.

séquence nucléotidique SEQ ID N°1 ou un fragment de celle-ci.

Un autre objet de la présente invention est un vecteur pour le clonage, l'expression et/ou l'insertion d'une  
5 séquence, caractérisé en ce qu'il comprend une séquence nucléotidique de polynucléotide selon l'invention en un site non essentiel pour sa répllication, le cas échéant sous le contrôle d'éléments de régulation susceptibles d'intervenir dans l'expression du polypeptide DP428, chez  
10 un hôte déterminé.

Des vecteurs particuliers sont par exemple des plasmides, des phages, des cosmides, des phagemides, des YAC.

Ces vecteurs sont utiles pour transformer des cellules  
15 hôtes afin de cloner ou d'exprimer les séquences nucléotidiques de l'invention.

L'invention comprend également les cellules hôtes transformées par un vecteur selon l'invention.

20

De préférence, les cellules hôtes sont transformées dans des conditions permettant l'expression d'un polypeptide recombinant selon l'invention.

25 Une cellule hôte préférée selon l'invention est la souche *E. coli* transformée par le plasmide pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°1-1818 ou transformée par le plasmide pM1025 déposé le 4 août 1998 à la CNCM sous le n° 1-2062 ou une mycobactérie appartenant à une souche  
30 de *M. tuberculosis*, *M. bovis* ou *M. africanum* possédant potentiellement tous les systèmes de régulation appropriés.

Il est aujourd'hui facile de produire des protéines

4. L'invention a pour objet une méthode pour obtenir un polypeptide recombinant à l'aide d'un vecteur d'expression approprié et d'un

produire *in vitro* un polypeptide selon l'invention, notamment le polypeptide DP428. Ledit polypeptide pourra être fixé sur une microplaque pour développer un test sérologique destiné à rechercher, dans un but de diagnostic, les anticorps spécifiques chez les patients atteints de tuberculose.

Ainsi, la présente invention concerne un procédé de préparation d'un polypeptide, caractérisé en ce qu'il met en oeuvre un vecteur selon l'invention. Plus particulièrement l'invention concerne un procédé de préparation d'un polypeptide de l'invention comprenant les étapes suivantes :

- le cas échéant, l'amplification préalable suivant la technique PCR de la quantité de séquences de nucléotides codant pour ledit polypeptide à l'aide de deux amorces d'ADN choisies de manière à ce que l'une de ces amorces soit identique aux 10 à 25 premiers nucléotides de la séquence nucléotidique codant pour ledit polypeptide, tandis que l'autre amorce est complémentaire des 10 à 25 derniers nucléotides (ou s'hybride avec ces 10 à 25 derniers nucléotides) de ladite séquence nucléotidique, ou inversement de manière à ce que l'une de ces amorces soit identique aux 10 à 25 derniers nucléotides de ladite séquence, tandis que l'autre amorce est complémentaire des 10 à 25 premiers nucléotides (ou s'hybride avec les 10 à 25 premiers nucléotides) de ladite séquence nucléotidique, suivie de l'insertion des dites séquences ainsi amplifiées dans un vecteur approprié,
- la mise en culture, dans un milieu de culture approprié, d'un hôte cellulaire préalablement transformé par un vecteur approprié contenant un acide nucléique selon l'invention.



L'invention a aussi pour objet un polypeptide susceptible d'être obtenu par un procédé de l'invention tel que décrit précédemment.

5 Les peptides selon l'invention peuvent également être préparés par les techniques classiques, dans le domaine de la synthèse des peptides. Cette synthèse peut être réalisée en solution homogène ou en phase solide.

10 Par exemple, on aura recours à la technique de synthèse en solution homogène décrite par Houbenweyl en 1974.

Cette méthode de synthèse consiste à condenser successivement deux-à-deux les aminoacyles successifs dans l'ordre requis, ou à condenser des aminoacyles et des 15 fragments préalablement formés et contenant déjà plusieurs aminoacyles dans l'ordre approprié, ou encore plusieurs fragments préalablement ainsi préparés, étant entendu que l'on aura eu soin de protéger au préalable toutes les 20 fonctions réactives portées par ces aminoacyles ou fragments, à l'exception des fonctions amines de l'un et carboxyles de l'autre ou vice versa, qui doivent normalement intervenir dans la formation des liaisons peptidiques, notamment après activation de la fonction 25 carboxyle, selon les méthodes bien connues dans la synthèse des peptides. En variante, on pourra avoir recours à des réactions de couplage mettant en jeu des réactifs de couplage classique, du type carbodiimide, tels que par exemple la 1-éthyl-3-(3-diméthylaminopropyl)-carbodiimide.

30 Lorsque l'aminoacyle en œuvre possède une fonction acide supplémentaire, notamment dans le cas de l'acide glutamique, ces fonctions doivent être protégées, par exemple par la formation d'esters.

désirée et ainsi de suite, de proche en proche, jusqu'à l'acide aminé N-terminal.

Selon une autre technique préférée de l'invention, on a recours à celle décrite par Merrifield.

5 Pour fabriquer une chaîne peptidique selon le procédé de Merrifield, on a recours à une résine polymère très poreuse, sur laquelle on fixe le premier acide aminé C-terminal de la chaîne. Cet acide aminé est fixé sur la résine par l'intermédiaire de son groupe carboxylique et sa  
10 fonction amine est protégée, par exemple par le groupe t-butyloxycarbonyle.

Lorsque le premier acide aminé C terminal est ainsi fixé sur la résine, on enlève le groupe protecteur de la  
15 fonction amine en lavant la résine avec un acide.

Dans le cas où le groupe protecteur de la fonction amine est le groupe t-butyloxycarbonyle, il peut être éliminé par traitement de la résine à l'aide d'acide  
20 trifluoroacétique.

On couple ensuite le deuxième acide aminé qui fournit le second aminoacyle de la séquence recherchée, à partir du résidu aminoacyle C-terminal sur la fonction amine  
25 déprotégée du premier acide aminé C-terminal fixé sur la chaîne. De préférence, la fonction carboxyle de ce deuxième acide aminé est activée, par exemple par la dicyclohexylcarbodiimide, et la fonction amine est protégée, par exemple par le t butyloxycarbonyle.

30 On obtient ainsi la première partie de la chaîne peptidique recherchée, qui comporte deux acides aminés, et dont la fonction amine terminale est protégée. Comme précédemment, on déprotège la fonction amine.

On fixe ainsi, les uns après les autres, les acides aminés qui vont constituer la chaîne peptidique sur le groupe amine chaque fois déprotégé au préalable de la portion de la chaîne peptidique déjà formée, et qui est  
5 rattachée à la résine.

Lorsque la totalité de la chaîne peptidique désirée est formée, on élimine les groupes protecteurs des différents acides aminés constituant la chaîne peptidique et on détache le peptide de la résine, par exemple à l'aide  
10 d'acide fluorhydrique.

De manière préférentielle, lesdits polypeptides susceptibles d'être obtenus par un procédé de l'invention tel que décrit précédemment comprendront une région exposée  
15 au solvant et auront une longueur d'au moins 20 acides aminés.

Selon un autre mode de réalisation de l'invention, lesdits polypeptides sont spécifiques de mycobactéries du complexe *Mycobacterium tuberculosis* et ne sont donc pas  
20 reconnus par des anticorps spécifiques d'autres protéines de mycobactéries.

L'invention est en outre relative à des polypeptides hybrides présentant au moins un polypeptide selon l'invention et une séquence d'un polypeptide susceptible  
25 d'induire une réponse immunitaire chez l'homme ou l'animal.

Avantageusement, le déterminant antigénique est tel qu'il est susceptible d'induire une réponse humorale et/ou  
30 cellulaire.

Un tel déterminant pourra comprendre un polypeptide

comportant une séquence d'au moins 10 acides aminés susceptible d'induire une réponse immunitaire.

Ces molécules hybrides peuvent être constituées en partie d'une molécule porteuse de polypeptide selon l'invention associée à une partie, en particulier un épitope de la toxine diphtérique, la toxine tétanique, un  
5 antigène de surface du virus de l'hépatite B (brevet FR 79 21811), l'antigène VP1 du virus de la poliomyélite ou toute autre toxine ou antigène viral ou bactérien.

Avantageusement, ledit déterminant antigénique  
10 correspond à un déterminant antigénique de protéines immunogènes de 45/47 kD de *M. tuberculosis* (demande internationale PCT/FR 96/0166), ou encore sélectionnées par exemple parmi ESAT6 (Harboe et al., 1996, Andersen et al., 1995, et Sorensen et al., 1995) et DES (PCT/FR 97/00923,  
15 Gicquel et al.).

Un antigène viral, tel que défini ci-dessus, sera préférentiellement une protéine de surface ou d'enveloppe d'un virus de l'hépatite, par exemple la protéine de  
20 surface de l'hépatite B sous l'une de ses formes S, S-prés1, S-prés2 ou S-prés2-prés1 ou encore une protéine d'un virus de l'hépatite A, ou d'une hépatite non-A, non-B, tel qu'un virus de l'hépatite C, E ou delta.

Plus particulièrement, un antigène viral tel que défini ci-dessus sera tout ou partie de l'une des glycoprotéines codées par le génome du virus HIV-1 (brevets EP 8324800, EP 84401834 ou EP 85905513), ou du virus HIV-2 (EP 87400151), et en particulier tout ou partie d'une  
30 protéine sélectionnée parmi gag, pol, nef ou env de HIV-1 ou de HIV-2.

d'obtention de gènes codant pour des protéines de fusion décrite par Minton en 1984.

Lesdits polynucléotides hybrides codant pour un polypeptide hybride ainsi que les polypeptides hybrides selon l'invention caractérisés en ce qu'il s'agit de protéines recombinantes obtenues par l'expression desdits polynucléotides hybrides, font également partie de l'invention.

10

Les polypeptides selon l'invention peuvent avantageusement être mis en oeuvre dans un procédé pour la détection *in vitro* d'anticorps dirigés contre lesdits polypeptides, notamment le polypeptide DP428, et ainsi d'anticorps dirigés contre une bactérie du complexe *Mycobacterium tuberculosis*, dans un échantillon biologique (tissu ou fluide biologique) susceptible de les contenir, ce procédé comprenant la mise en contact de cet échantillon biologique avec un polypeptide selon l'invention dans des conditions permettant une réaction immunoologique *in vitro* entre ledit polypeptide et les anticorps éventuellement présents dans l'échantillon biologique, et la mise en évidence *in vitro* des complexes antigène-anticorps éventuellement formés.

15

Les polypeptides selon l'invention peuvent également et avantageusement être mis en oeuvre dans un procédé pour la détection d'une infection par une bactérie du complexe *Mycobacterium tuberculosis* dans un mammifère basé sur la détection *in vitro* d'une réaction cellulaire indiquant une sensibilisation préalable du mammifère audit polypeptide.

30

- a) préparation d'un échantillon biologique contenant des cellules dudit mammifère plus particulièrement des cellules du système immunitaire dudit mammifère et plus particulièrement encore des cellules T ;
- 5 b) incubation de l'échantillon biologique de l'étape a) avec un polypeptide selon l'invention;
- c) détection d'une réaction cellulaire indiquant une sensibilisation préalable du mammifère audit polypeptide comme par exemple la prolifération cellulaire et/ou la
- 10 synthèse de protéines telles que l'interféron gamma.
- La prolifération cellulaire pourra être mesurée, par exemple par incorporation de <sup>3</sup>H-Thymidine.

Font également partie de l'invention, les procédés de

15 détection d'une réaction d'hypersensibilité retardée (DTH), caractérisés en ce qu'ils mettent en oeuvre un polypeptide selon l'invention.

De préférence, l'échantillon biologique est constitué

20 par un fluide, par exemple un sérum humain ou animal, du sang, des biopsies, le liquide broncho-alvéolaire ou le liquide pleural.

Toute procédure classique peut être mise en oeuvre

25 pour réaliser une telle détection.

A titre d'exemple, une méthode préférée met en jeu des processus immunoenzymatiques selon la technique ELISA, par immunofluorescence, ou radio immunologique (RIA) ou

30 équivalent.

Ainsi, l'invention concerne également les polypeptides

- dépôt de quantités déterminées d'une composition polypeptidique selon l'invention dans les puits d'une plaque de microtitration,
- introduction dans lesdits puits de dilutions croissantes de sérum, ou d'échantillon biologique autre tel que défini précédemment, devant être analysé,
- incubation de la microplaque,
- introduction dans les puits de la plaque de microtitration d'anticorps marqués dirigés contre des immunoglobulines humaines ou animales, le marquage de ces anticorps ayant été réalisé à l'aide d'une enzyme sélectionnée parmi celles qui sont capables d'hydrolyser un substrat en modifiant l'absorption des radiations de ce dernier, au moins à une longueur d'onde déterminée, par exemple à 550 nm,
- détection, en comparaison avec un témoin de contrôle, de la quantité de substrat hydrolysé.

L'invention concerne également un nécessaire ou kit pour le diagnostic *in vitro* d'une infection par une mycobactérie appartenant au complexe *Mycobacterium tuberculosis*, comprenant:

- un polypeptide selon l'invention,
- le cas échéant les réactifs pour la constitution du milieu propice à la réaction immunologique ou spécifique,
- les réactifs permettant la détection des complexes antigène anticorps produits par la réaction immunologique éventuellement présente dans l'échantillon biologique, et la mise en évidence *in vitro* des complexes antigène-anticorps éventuellement formés, ces réactifs pouvant également porter un marqueur, ou être susceptibles d'être reconnus à leur tour par un réactif marqué.

- le cas échéant, un échantillon biologique de référence (témoin positif) contenant une quantité prédéterminée d'anticorps reconnus par un polypeptide selon l'invention.

5 Les polypeptides selon l'invention permettent de préparer des anticorps monoclonaux ou polyclonaux caractérisés en ce qu'ils reconnaissent spécifiquement les polypeptides selon l'invention. Les anticorps monoclonaux  
10 pourront avantageusement être préparés à partir d'hybridomes selon la technique décrite par Kohler et Milstein en 1975. Les anticorps polyclonaux pourront être préparés, par exemple par immunisation d'un animal, en particulier une souris, avec un polypeptide selon l'invention associé à un adjuvant de la réponse  
15 immunitaire, puis purification des anticorps spécifiques contenus dans le sérum des animaux immunisés sur une colonne d'affinité sur laquelle a préalablement été fixé le polypeptide ayant servi d'antigène. Les anticorps polyclonaux selon l'invention peuvent aussi être préparés  
20 par purification sur une colonne d'affinité, sur laquelle a préalablement été immobilisé un polypeptide selon l'invention, des anticorps contenus dans le sérum de patients infectés par une mycobactérie et préférentiellement une bactérie appartenant au complexe  
25 *Mycobacterium tuberculosis*.

L'invention a également pour objet des anticorps mono  
ou polyclonaux ou leurs fragments, ou anticorps  
30 chimériques, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'invention.



sondes nucléiques de l'invention tel qu'un marquage de type enzymatique, fluorescent ou radioactif.

L'invention vise en outre un procédé pour la détection  
5 spécifique de la présence d'un antigène d'une mycobactérie et préférentiellement une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- 10 a) Mise en contact de l'échantillon biologique (tissu ou fluide biologique) prélevé chez un individu avec un anticorps mono ou polyclonal selon l'invention, dans des conditions permettant une réaction immunologique in vitro entre lesdits anticorps et les polypeptides spécifiques des  
15 mycobactéries et préférentiellement des bactéries du complexe de *Mycobacterium tuberculosis* éventuellement présents dans l'échantillon biologique, et
- b) Mise en évidence du complexe antigène-anticorps formé.

20 Entre également dans le cadre de l'invention, un nécessaire ou kit pour le diagnostic in vitro sur un échantillon biologique, de la présence de souches de mycobactéries des mycobactéries et préférentiellement des bactéries appartenant au complexe de *Mycobacterium tuberculosis*, de préférence *M. tuberculosis*, caractérisé en  
25 ce qu'il comprend :

- un anticorps polyclonal ou monoclonal selon l'invention, le cas échéant marqué;
- le cas échéant, un réactif pour la constitution du milieu propice à la réalisation de la réaction immunologique;
- 30 - un réactif permettant la détection des complexes antigène-anticorps.

- le cas échéant, des réactifs pour effectuer la lyse des cellules de l'échantillon testé.

La présente invention a également pour objet un  
5 procédé de détection et d'identification rapide des mycobactéries et préférentiellement des bactéries de *M. tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comporte les étapes suivantes :

- 10 a) Isolement de l'ADN à partir de l'échantillon biologique à analyser, ou obtention d'un ADNc à partir de l'ARN de l'échantillon biologique ;
- b) Amplification spécifique de l'ADN des mycobactéries et préférentiellement des bactéries appartenant au complexe *Mycobacterium tuberculosis* à l'aide d'amorces selon  
15 l'invention;
- c) Analyse des produits d'amplification.

Le produits d'amplification peuvent être analysés par différentes méthodes.

20 Deux méthodes d'analyse sont données à titre d'exemple ci-dessous :

- Analyse électrophorétique en gel d'agarose des produits d'amplification. La présence d'un fragment d'ADN migrant à l'endroit attendu suggère que l'échantillon  
25 analysé contenait de l'ADN de mycobactéries appartenant au complexe *tuberculosis*, ou

- Analyse par la technique d'hybridation moléculaire en utilisant une sonde nucléique selon l'invention. Cette sonde sera avantageusement marquée par un élément non  
30 radioactif (sonde froide) ou radioactif.

Aux fins de la présente invention, on entend par :

l'ADN : tout matériel génétique, qu'il soit isolé ou non, et qu'il soit présent sous forme d'acide nucléique pur ou sous forme de complexe avec des protéines ou des lipides.

Un autre procédé de la présente invention permet la détection d'une infection par une mycobactérie et  
5 préférentiellement une bactérie du complexe *Mycobacterium tuberculosis* dans un mammifère. Ce procédé comprend les étapes suivantes :

- a) préparation d'un échantillon biologique contenant des  
10 cellules dudit mammifère plus particulièrement des cellules du système immunitaire dudit mammifère et plus particulièrement encore des cellules T ;
- b) incubation de l'échantillon biologique de l'étape a) avec un polypeptide selon l'invention;
- c) détection d'une réaction cellulaire indiquant une  
15 sensibilisation préalable du mammifère audit polypeptide, notamment la prolifération cellulaire et/ou la synthèse de protéines telles que l'interféron gamma;
- d) détection d'une réaction d'hypersensibilité retardée ou de sensibilisation du mammifère audit polypeptide.

20 Cette méthode de détection est une méthode intradermique, qui est décrite par exemple par M. J. Elhay et al. (1988) *Infection and Immunity*, 66(7) : 3454-3456.

25 Un autre but de la présente invention consiste en un procédé pour la détection des mycobactéries et préférentiellement des bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- 30 a) Mise en contact d'une sonde oligonucléotidique selon l'invention avec un échantillon biologique, l'ADN contenu dans l'échantillon biologique, ou l'ARN.

1. Procédé de détection d'une infection par une mycobactérie et préférentiellement une bactérie du complexe *Mycobacterium tuberculosis* dans un mammifère, caractérisé en ce qu'il comprend les étapes suivantes :

préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*;

b) Détection de l'hybride formé entre la sonde oligonucléotidique et l'ADN de l'échantillon biologique.

5

L'invention vise également un procédé pour la détection des mycobactéries et préférentiellement des bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend

10 les étapes suivantes :

a) Mise en contact d'une sonde oligonucléotidique selon l'invention immobilisée sur un support, avec un échantillon biologique, l'ADN de l'échantillon biologique ayant, le cas échéant, été préalablement rendu accessible à

15 l'hybridation, dans des conditions permettant l'hybridation de ladite sonde à l'ADN des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*;

b) Mise en contact de l'hybride formé entre ladite sonde oligonucléotidique immobilisée sur un support et l'ADN contenu dans l'échantillon biologique, le cas échéant après élimination de l'ADN de l'échantillon biologique n'ayant pas hybridé avec la sonde, avec une sonde

20 25

Selon un mode de réalisation avantageux du procédé de détection défini précédemment, celui-ci est caractérisé en ce que, préalablement à l'étape a), l'ADN de l'échantillon biologique est préalablement amplifié à l'aide d'un couple d'amorces selon l'invention.

30

Une autre forme de mise en œuvre de l'invention

*Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

5 a) Mise en contact de l'échantillon biologique avec un couple d'amorces selon l'invention, l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des conditions permettant une hybridation desdites amorces à l'ADN des mycobactéries et préférentiellement des bactéries du  
10 complexe *Mycobacterium tuberculosis*;

b) Amplification de l'ADN d'une mycobactérie et préférentiellement d'une bactérie du complexe *Mycobacterium tuberculosis*;

15 c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par les amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique selon invention.

20 L'invention a aussi pour objet un procédé pour la détection de la présence des mycobactéries et préférentiellement des bactéries appartenant au complexe de *Mycobacterium tuberculosis* dans un échantillon biologique par déplacement de brin, caractérisé en ce qu'il comprend les étapes suivantes :

25 a) Mise en contact de l'échantillon biologique avec deux couples d'amorces selon l'invention spécifiquement destinées à l'amplification de type SPA décrites ci-dessus, l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des  
30 conditions permettant une hybridation.

- b) amplification de l'ADN des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*;
- c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par les amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique selon l'invention.

L'invention concerne aussi un nécessaire ou kit pour la mise en oeuvre du procédé décrit ci-dessus, destiné à la détection de la présence des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

- a) Une sonde oligonucléotidique selon l'invention;
- b) Les réactifs nécessaires à la mise en oeuvre d'une réaction d'hybridation;
- c) Le cas échéant, un couple d'amorces selon l'invention ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN (ADN génomique, ADN plasmidique ou ADNc) des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*.

L'invention a aussi pour objet un kit ou nécessaire pour la détection de la présence des mycobactéries et préférentiellement des bactéries du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :

- a) Une sonde oligonucléotidique, dite sonde de capture, selon l'invention;
- b) Une sonde oligonucléotidique, dite sonde de

préférentiellement des bactéries du complexe *Mycobacterium tuberculosis*.

L'invention concerne encore un kit ou nécessaire pour  
5 l'amplification de l'ADN des mycobactéries et  
préférentiellement des bactéries du complexe *Mycobacterium tuberculosis* présent dans un échantillon biologique,  
caractérisé en ce qu'il comprend les éléments suivants :

- 10 a) Un couple d'amorces selon l'invention;
- b) Les réactifs nécessaires pour effectuer une réaction  
d'amplification d'ADN;
- c) Eventuellement un composant permettant de vérifier la  
séquence du fragment amplifié, plus particulièrement une  
15 sonde oligonucléotidique selon l'invention.

Un autre objet de la présente invention concerne une  
composition immunogène, caractérisée en ce qu'elle comprend  
un polypeptide selon l'invention.

20 Une autre composition immunogène selon l'invention est  
caractérisée en ce qu'elle comprend un ou plusieurs  
polypeptides selon l'invention et/ou un ou plusieurs  
polypeptides hybrides selon l'invention.

25 Selon un mode de réalisation avantageux, la  
composition immunogène ci-dessus définie est constitutive  
d'un vaccin, lorsqu'elle est présentée en association avec  
un véhicule pharmaceutiquement acceptable et éventuellement  
un ou plusieurs adjuvants de l'immunité tels que l'alun ou  
un représentant de la famille des squalyl peptides ou  
encore l'adjuvant incomplet de Freund.

tuberculose), micro organismes inactivés (virus de la grippe), des extraits acellulaires (*Bordetella pertussis* pour la coqueluche), protéines recombinées (antigène de surface du virus de l'hépatite B), des polysides (pneumocoques). Des vaccins préparés à partir de peptides de synthèse ou de micro-organismes génétiquement modifiés exprimant des antigènes hétérologues sont en cours d'expérimentation. Plus récemment encore, des ADN plasmidiques recombinés portant des gènes codant pour des antigènes protecteurs ont été proposés comme stratégie vaccinale alternative. Ce type de vaccination est réalisé avec un plasmide particulier dérivant d'un plasmide de *E. coli* qui ne se réplique pas *in vivo* et qui code uniquement pour la protéine vaccinante. Les principaux composants fonctionnels de ce plasmide sont : un promoteur fort permettant l'expression dans les cellules eucaryotes (par exemple celui du CMV), un site de clonage approprié pour insérer le gène d'intérêt, une séquence de terminaison-polyadénylation, une origine de réplication procaryote pour produire le plasmide recombiné *in vitro* et un marqueur de sélection (par exemple le gène de résistance à l'ampicilline) pour faciliter la sélection des bactéries qui contiennent le plasmide. Des animaux ont été immunisés en injectant simplement l'ADN plasmidique nu dans le muscle. Cette technique conduit à l'expression de la protéine vaccinale *in situ* et à une réponse immunitaire en particulier de type cellulaire (CTL) et de type humoral (anticorps). Cette double induction de la réponse immunitaire est l'un des principaux avantages de la technique de vaccination avec de l'ADN nu. Huygen et al. (1996) et Tascón et al. (1996) ont réussi à obtenir une certaine protection.

Il est également possible d'insérer dans le plasmide des gènes codant pour des protéines adjuvantes, par exemple.



tout ou partie du gène DP428, objet de la présente invention conduirait probablement à une meilleure protection contre la tuberculose. Tout ou partie du gène DP428, ou tout polynucléotide selon l'invention, peut être  
5 facilement inséré dans les plasmides vecteurs VIJ (Montgomery et al, 1993), pcDNA3 (Invitrogen, R & D Systems) ou pcDNA1/Neo (Invitrogen) qui possèdent les caractéristiques nécessaires pour une utilisation vaccinale.

10

L'invention vise ainsi un vaccin, caractérisée en ce qu'il comprend un ou plusieurs polypeptides selon l'invention et/ou un ou plusieurs polypeptides hybrides selon l'invention tels que précédemment définis en  
15 association avec un véhicule pharmaceutiquement compatible et, le cas échéant, un ou plusieurs adjuvants de l'immunité appropriés.

L'invention vise aussi une composition vaccinale destinée à l'immunisation de l'homme ou l'animal à l'encontre d'une infection bactérienne ou virale, telle que la tuberculose ou l'hépatite, caractérisée en ce qu'elle comprend un ou plusieurs polypeptides hybrides tels que précédemment définis en association avec un véhicule  
20 pharmaceutiquement compatible et, le cas échéant, un ou plusieurs adjuvants de l'immunité.

Avantageusement, dans le cas d'une protéine hybride entre un polypeptide selon l'invention et l'antigène de surface de l'hépatite B, la composition vaccinale sera administrée, chez l'homme, à raison de 0,1 à 1 µg de protéine hybride purifiée par kilogramme du poids du patient.

30

On peut également envisager d'administrer, de préférence, la

moitié de la quantité pondérale de la protéine hybride contenue dans une dose destinée à un patient n'étant pas affecté de troubles du système immunitaire.

5 De préférence, la composition vaccinale sera administrée à plusieurs reprises, de manière étalée dans le temps, par voie intradermique ou sous-cutanée. A titre d'exemple, trois doses telles que définies ci-dessus seront respectivement administrées au patient au temps  $t_0$ , au  
10 temps  $t_0 + 1$  mois et au temps  $t_0 + 1$  an.

Alternativement, trois doses seront respectivement administrées au patient au temps  $t_0$ , au temps  $t_0 + 1$  mois et au temps  $t_0 + 6$  mois.

15 Chez la souris, chez laquelle une dose pondérale de la composition vaccinale comparable à la dose utilisée chez l'homme est administrée, la réaction anticorps est testée par prélèvement du sérum suivi d'une étude de la formation d'un complexe entre les anticorps présents dans le sérum et  
20 l'antigène de la composition vaccinale, selon les techniques usuelles.

● L'invention concerne également une composition immunogène caractérisée en ce qu'elle comprend un  
25 polynucléotide ou un vecteur d'expression selon l'invention, en association avec un véhicule permettant son administration à l'homme ou l'animal.

L'invention a encore pour objet un vaccin destiné à  
30 l'immunisation à l'encontre d'une infection bactérienne ou virale, telle que la tuberculose ou l'hépatite, caractérisé en ce qu'il comprend un polynucléotide ou un vecteur d'expression selon l'invention.

90/11092 (Vical Inc.) et également dans la demande internationale N° WO 95/11307 (Institut Pasteur).

Le polynucléotide constitutif de la composition  
5 immunogène ou de la composition vaccinale selon l'invention  
peut être injecté à l'hôte après avoir été couplé à des  
composés qui favorisent la pénétration de ce polynucléotide  
à l'intérieur de la cellule ou son transport jusqu'au noyau  
cellulaire. Les conjugués résultants peuvent être  
10 encapsulés dans des microparticules polymères, comme décrit  
dans la demande internationale N° WO 94/27238 (medisorb  
Technologies International).

Selon un autre mode de réalisation de la composition  
15 immunogène et/ou vaccinale selon l'invention, le  
polynucléotide, de préférence un ADN, est complexé avec du  
DEAE-dextran (Pagano et al., 1967) ou avec des protéines  
nucléaires (Kaneda et al., 1989), avec des lipides (Felgner  
et al., 1987) ou encore encapsulés dans des liposomes  
20 (Fraley et al., 1980).

Selon encore un autre mode de réalisation avantageux  
de la composition immunogène et/ou vaccinale selon  
l'invention, le polynucléotide selon l'invention peut être  
25 introduit sous la forme d'un gel facilitant sa transfection  
dans les cellules. Une telle composition sous forme de gel  
peut être un complexe de poly-L-lysine et de lactose, comme  
décrit par Midoux en 1993, ou encore le Poloxamer 407<sup>®</sup>,  
comme décrit par Pastore en 1994. Le polynucléotide ou le  
30 vecteur selon l'invention peuvent aussi être en suspension  
dans une solution tampon ou être associés à des liposomes.

Avantageusement, un tel

Un tel vaccin sera avantageusement préparé sous la forme d'une composition contenant un vecteur selon l'invention, placée sous le contrôle d'éléments de régulation permettant son expression chez l'homme ou  
5 l'animal.

Pour réaliser un tel vaccin, le polynucléotide selon l'invention est tout d'abord sous-cloné dans un vecteur d'expression approprié, plus particulièrement un vecteur  
10 d'expression contenant des signaux de régulation et d'expression reconnus par les enzymes des cellules eucaryotes et contenant également une origine de réplication active chez les procaryotes, par exemple chez *E. coli*, qui permet son amplification préalable. Puis le  
15 plasmide recombinant purifié obtenu est injecté à l'hôte, par exemple par voie intramusculaire.

On pourra par exemple utiliser, en tant que vecteur d'expression in vivo de l'antigène d'intérêt, le plasmide pcDNA3 ou le plasmide pcDNA1/neo, tous les deux  
20 commercialisés par Invitrogen (R&D Systems, Abingdon, Royaume-Uni). On peut aussi utiliser le plasmide V1Jns.tPA, décrit par Shiver et al. en 1995.

Un tel vaccin comprendra avantageusement, outre le vecteur recombinant, une solution saline, par exemple une  
25 solution de chlorure de sodium.

Une composition vaccinale telle que définie ci-dessus sera par exemple administrée par voie parentérale ou par voie intramusculaire.

30

La présente invention concerne également un vaccin caractérisé en ce qu'il comprend :

échéant, un ou plusieurs adjuvants de l'immunité appropriés.

Un autre aspect porte sur une méthode de criblage de  
5 molécules capables d'inhiber la croissance de mycobactéries  
ou le maintien de mycobactéries dans un hôte, caractérisée  
en ce que lesdites molécules bloquent la synthèse ou la  
fonction des polypeptides codés par une séquence  
nucléotidique selon l'invention ou par un polynucléotide  
10 tel que décrit supra.

Dans ladite méthode de criblage, les molécules peuvent  
être des anti-messagers ou peuvent induire la synthèse  
d'anti-messagers.

15

La présente invention vise également des molécules  
capables d'inhiber la croissance de mycobactéries ou le  
maintien de mycobactéries dans un hôte, caractérisées en ce  
que lesdites molécules sont synthétisées d'après la  
20 structure des polypeptides codés par une séquence  
nucléotidique selon l'invention ou par un polynucléotide  
tel que décrit supra.

D'autres caractéristiques et avantages de l'invention  
25 apparaissent dans les exemples et les figures suivants :

#### FIGURES

30 La série de Figures 1 :

FIGURE 1 : Schéma de la structure d'un polypeptide codé par une séquence nucléotidique selon l'invention.

polypeptides codés par la série des séquences nucléotidiques SEQ ID N°1.

Figure 2 :

- 5 Illustre la séquence nucléotidique SEQ ID N°2 correspondant à la région incluant le gène codant pour le polypeptide DP428 (région soulignée). Sur cette figure ont été pris en compte à la fois les codons ATG et CTG d'initiation de la traduction. La figure fait  
10 apparaître que le polypeptide DP428 fait probablement partie d'un opéron comprenant au moins trois gènes. La région doublement encadrée inclut probablement les régions promotrices.
- 15 La région simplement encadrée correspond au motif LPISG rappelant le motif LPXTG décrit chez les bactéries à Gram positifs comme permettant l'ancrage aux peptidoglycannes.

20 La série de Figures 3 :

La série de Figures 3 représente la série de séquences nucléotidiques SEQ ID N°3 correspondant à l'insert du vecteur p6D7 (déposé à la CNCM sous le N° I-1814).

25

La série de Figures 4 :

La série de Figures 4 représente la série de séquences nucléotidiques SEQ ID N°4 correspondant à l'insert du vecteur p5A3 (déposé à la CNCM sous le N° I-1815).

30

La série de Figures 5 :

La série de Figures 6 :

La série de Figures 6 représente la série de séquences  
nucléotidiques SEQ ID N°6 correspondant à l'insert du  
5 vecteur p2A29 (déposé à la CNCM sous le N° I-1817).

La série de Figures 7 :

La série de Figures 7 représente la série de séquences  
10 nucléotidiques SEQ ID N°7 correspondant à l'insert du  
vecteur p5B5 (déposé à la CNCM sous le N° I-1819).

La série de Figures 8 :

La série de Figures 8 représente série de séquences  
15 nucléotidiques SEQ ID N°8 correspondant à l'insert du  
vecteur p1C7 (déposé à la CNCM sous le N° I-1820).

La série de Figures 9 :

20 La série de Figures 9 représente la série de séquences  
nucléotidiques SEQ ID N°9 correspondant à l'insert du  
vecteur p2D7 (déposé à la CNCM sous le N° I-1821).

La série de Figures 10 :

25 La série de Figures 10 représente la série de séquences  
nucléotidiques SEQ ID N°10 correspondant à l'insert du  
vecteur p1B7 (déposé à la CNCM sous le N° I-1843).

30 La série de Figures 11 :

La série de Figures 11 représente la série de séquences  
nucléotidiques SEQ ID N°11 correspondant à l'insert du  
vecteur p1B7 (déposé à la CNCM sous le N° I-1843).

La série de Figures 12 représente la série de séquences nucléotidiques SEQ ID N°12.

La série de Figures 13 :

5

La série de Figures 13 représente la série de séquences nucléotidiques SEQ ID N°13.

La série de Figures 14 :

10

La série de Figures 14 représente la série de séquences nucléotidiques SEQ ID N°14 correspondant à l'insert du vecteur p5B5 (déposé à la CNCM sous le N° I-1819).

15

La série de Figures 15 :

La série de Figures 15 représente la série de séquences nucléotidiques SEQ ID N°15.

20

La série de Figures 16 :

La série de Figures 16 représente la série de séquences nucléotidiques SEQ ID N°16.

25

La série de Figures 17 :

La série de Figures 17 représente la série de séquences nucléotidiques SEQ ID N°17.

30

La série de Figures 18 :

La série de Figures 18 :



La série de Figures 19 représente la série de séquences nucléotidiques SEQ ID N°19.

5 La série de Figures 20 :

La série de Figures 20 représente la série de séquences nucléotidiques SEQ ID N°20 correspondant à l'insert du vecteur p2A29 (déposé à la CNCM sous le N° I-1817).

10

La série de Figures 21 :

La série de Figures 21 représente la série de séquences nucléotidiques SEQ ID N°21.

15

La série de Figures 22 :

La série de Figures 22 représente la série de séquences nucléotidiques SEQ ID N°22.

20

La série de Figures 23 :

La série de Figures 23 représente la série de séquences nucléotidiques SEQ ID N°23.

25

La série de Figures 24 :

La série de Figures 24 représente la série de séquences nucléotidiques SEQ ID N°24.

30

Figures 25 et 26 :

Les figures 25 et 26 illustrent respectivement les séquences SEQ ID N°25 et SEQ ID N°26.

La série de Figures 27 :

La série de Figures 27 représente la série de séquences  
5 nucléotidiques SEQ ID N°27 correspondant à l'insert du  
vecteur p5A3.

Figure 28 :

10 La séquence d'acides aminés telle que définie dans la  
figure 28 représente la séquence d'acides aminés SEQ ID  
N°28 correspondant au polypeptide DP428.

15 Figure 29 :

La figure 29 représente la séquence nucléotidique SEQ ID N°  
29 du gène complet codant pour la protéine MLC25.

20 Figure 30 :

La figure 30 représente la séquence d'acides aminés SEQ ID  
N° 30 de la protéine MLC25.

25 La série de Figures 31 :

La série de Figures 31 représente la série de séquences  
nucléotidiques SEQ ID N° 31.

30 La série de Figures 32 :

La série de Figures 32 représente la série de séquences  
nucléotidiques SEQ ID N° 32.

La série de Figures 33 représente la série de séquences nucléotidiques SEQ ID N°33.

La série de Figures 34 :

5

La série de Figures 32 représente la série de séquences nucléotidiques SEQ ID N°34.

La série de Figures 35 :

10

La série de Figures 35 représente la série de séquences nucléotidiques SEQ ID N°35.

15

La série de Figures 36 :

La série de Figures 36 représente la série de séquences nucléotidiques SEQ ID N°36.

20

La série de Figures 37 :

La série de Figures 37 représente la série de séquences nucléotidiques SEQ ID N°37.

25

La série de Figures 38 :

La série de Figures 38 représente la série de séquences nucléotidiques SEQ ID N°38.

30

La série de Figures 39 :

La série de Figures 39 représente la série de séquences nucléotidiques SEQ ID N°39.

La série de Figures 40 représente la série de séquences nucléotidiques SEQ ID N°40.

La série de Figures 41 :

5

La série de Figures 41 représente la série de séquences nucléotidiques SEQ ID N°41 correspondant à l'insert du vecteur p2D7 (déposé à la CNCM sous le N°I-1821).

10 La série de Figures 42 :

La série de Figures 42 représente la série de séquences nucléotidiques SEQ ID N°42.

15

La série de Figures 43 :

20 La série de Figures 43 représente la série de séquences nucléotidiques SEQ ID N°43.

La série de Figures 44 :

25 La série de Figures 44 représente la série de séquences nucléotidiques SEQ ID N°44.

La série de Figures 45 :

30 La série de Figures 45 représente la série de séquences nucléotidiques SEQ ID N°45.

La série de Figures 46 :

La série de Figures 47 représente la série de séquences nucléotidiques SEQ ID N°47.

5    La série de Figures 48 :

La série de Figures 48 représente la série de séquences nucléotidiques SEQ ID N°48.

10   La série de Figures 49 :

La série de Figures 49 représente la série de séquences nucléotidiques SEQ ID N°49.

15

La série de Figures 50 :

La série de Figures 50 représente la série de séquences nucléotidiques SEQ ID N°50.

20

Figure 51 :

25    A.    la construction pJVED: Plasmid navette( pouvant se multiplier chez les mycobactéries ainsi que chez *E.coli*), avec un gène de résistance à la kanamycine (issu de Tn903) comme marqueur de sélection. Le gène *phoA* tronqué (*A phoA*) et le gène *luc* forment un opéron synthétique.

B.    Séquence de la jonction entre *phoA* et *luc*.

30

Figure 51 :

position nt 964 (extrémité 5' de la sonde) et le nucléotide en position nt 1234 (extrémité 3' de la sonde), extrémités incluses, de la séquence SEQ ID N°1.

5 Figures 53 et 54 :

Activités Luc et PhoA de *M. smegmatis* recombinant contenant le pJVED avec différents fragments nucléotidiques comme décrits en exemple. Les figures 52 et 53 représentent les résultats obtenus pour deux expériences distinctes réalisées dans les mêmes conditions.

15

Figure 55 :

Représentation de l'hydrophobicité (Kyte et Doolittle) de la séquence codante du polypeptide DP428 avec sa représentation schématisque. Le motif LPISG précède immédiatement la région C-terminale hydrophobe. La séquence se termine par deux arginines.

Figure 56 :

25 Représentation de l'hydrophobicité (Kyte et Doolittle) de la séquence du polypeptide M1028 de séquence d'acides aminés SEQ ID N° 30.

Figure 57 :

30

A- Gel d'acrylamide (12%) en condition dénaturante d'un extrait cellulaire obtenu par lysis de bactéries.

ligne 1: Marqueur de masse molaire (Prestained SDS-PAGE Standards High Range BIO-RAD®).

5 ligne 2: Extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 sans induction par l'IPTG.

10 ligne 3: Extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 après 4 heures d'induction par l'IPTG.

ligne 4: Marqueur de masse molaire (Prestained SDS-PAGE Standards Low Range BIO-RAD®).

15 B- Western blot d'un gel semblable gel (acrylamide 12%) révélé grâce à l'anticorps penta-His commercialisé par la société Quiagen.

20 ligne 1: représentation du marqueur de masse molaire (Prestained SDS-PAGE Standards High Range BIO-RAD®).

25 ligne 2: extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 sans induction par l'IPTG.

ligne 3: extrait bactérien obtenu par sonication de bactéries *E. coli* M15 contenant le plasmide pM1C25 après 4 heures d'induction par l'IPTG.

30 ligne 4: représentation du marqueur de masse molaire (Prestained SDS-PAGE Standards Low Range BIO-RAD®).

28400 daltons, correspond à l'expression de l'insert M1C25 cloné dans le vecteur pQE-60 (Qiagen®).

5        En ce qui concerne les légendes des autres figures qui sont numérotées par un caractère alphanumérique, chacune de ces autres figures représente la séquence nucléotidique et la séquence d'acides aminés de séquence SEQ ID dont la numérotation est identique au caractère alphanumérique de  
10        chacune desdites figures.

Les numérotations alphanumériques des figures représentant les SEQ ID comportant un nombre suivi d'une lettre ont les significations suivantes :

- les numérotations alphanumériques présentant le même  
15        nombre concernent une même famille de séquence rattachées à la séquence de référence SEQ ID dont la numérotation présente ce même nombre et la lettre A ;
- les lettres A, B et C pour une même famille de séquences distinguent les trois phases de lecture possibles de la  
20        séquence nucléotidique SEQ ID de référence (A) ;
- les lettres indexées par un prime (') signifient que la séquence correspond à un fragment de la séquence SEQ ID de  
25        référence (A) ;
- la lettre D signifie que la séquence correspond à la séquence du gène prédit par Cole et al., 1998 ;
- la lettre F signifie que la séquence correspond à la phase ouverte de lecture ORF pour "Open Reading Frame" contenant la séquence "P" correspondante d'après Cole et  
30        al., 1998 ;
- la lettre G signifie que la séquence est une séquence prédite par Cole et al., 1998, et présentant une homologie de plus de 70% avec la séquence SEQ ID de référence (A) ;



séquence "D" correspondante et pouvant être en phase avec la séquence "D" en raison d'erreurs de séquençage possibles ;

- la lettre P signifie que la séquence correspond à la phase ouverte de lecture contenant la séquence "R" correspondante ;

- la lettre Q signifie que la séquence correspond à une séquence contenant les séquences "F" et "P" correspondantes.

En ce qui concerne la famille de séquences SEQ ID N° 4, l'insert précédent *phoA* contient deux fragments non contigus sur le génome, SEQ ID 4J et SEQ ID 4A, et donc issus d'un clonage multiple permettant l'expression et l'exportation de *phoA*. Ces deux fragments non contigus, les gènes et les phases ouvertes de lecture qui les contiennent d'après Cole et al., 1998, sont importants pour l'exportation d'un polypeptide antigène ;

- les lettres J, K et L distinguent les trois phases de lecture possibles de la séquence nucléotidique "J" correspondante ;

- la lettre M signifie que la séquence correspond à la séquence prédite par Cole et al., 1998, et contenant la séquence SEQ ID N° 4J ;

- la lettre N signifie que la séquence correspond à la phase ouverte de lecture contenant la séquence SEQ ID N° 4M.

En ce qui concerne la famille de séquences SEQ ID N° 45, la lettre Z signifie que la séquence correspond à la séquence d'un fragment cloné fusionné avec *phoA*.

Enfin, en ce qui concerne la famille de séquence SEQ ID N° 41, la lettre S signifie que la séquence correspond à une séquence prédite par Cole et al., 1998 et fusionnée avec *phoA*.

### EXEMPLES

#### Matériel et méthodes

#### 5 Cultures bactériennes, plasmides et milieux de cultures

*E. coli* a été cultivé sur milieu liquide ou solide Luria-Bertani (LB). *M. smegmatis* a été cultivé sur milieu liquide Middlebrook 7H9 (Difco) additionné de dextrose albumine (ADC), 0,2 % de glycérol et 0,05 % de Tween, ou sur milieu  
10 solide L. Si nécessaire, l'antibiotique kanamycine a été rajouté à une concentration de 20 µg/ml-1. Les clones bactériens présentant une activité PhoA ont été détectés sur de l'agar LB contenant du 5-bromo 4-chloro-3-indolyile  
15 phosphate (X-P, à 40 µg/ml-1).

#### Manipulation d'ADN et séquençage

Les manipulations d'ADN et les analyses par  
20 Southern blot ont été effectuées en utilisant les techniques standard (Sambrook et al., 1989). Les séquences d'ADN double brin ont été déterminées avec un kit de séquençage Taq Dye Deoxy Terminator Cycle (Applied Biosystems), dans un Système 9600 GeneAmp PCR (Perkin-  
25 Elmer), et après migration sur un système d'analyse ADN modèle 373 (Applied Biosystems).

#### Constructions des plasmides

30 Le plasmide pJVEB<sub>1</sub> a été construit à partir de pLA71, plasmide de transfert comportant le gène *phoA* tronqué et placé en phase de lecture.

pGEM-luc et un site de liaison du ribosome a été rajouté. *phoA* a été amplifié à partir de pJEM11. Les fragments amplifiés ont été coupés avec *Pst*I et ligaturés ensemble. Les oligodéoxynucléotides utilisés sont les suivants :

- 5 pPV.luc.Fw : 5'GACTGCTGCAGAAGGAGAAGATCCAAATGG3'  
 luc.Bw : 5'GACTAGCGGCCGCGAATTCGTCCGACCTCCGAGG3'  
 pJEM.phoA.Fw : 5'CCGCGGATCCGGATACGTAC3'  
 phoA.Bw: 5'GACTGCTGCAGTTTATTTCAGCCCCAGAGCG3'.

Le fragment ainsi obtenu a été réamplifié en  
 10 utilisant les oligonucléotides complémentaires de ses extrémités, coupé avec *Kpn*I et *Not*I, et intégré dans pLA71 coupé avec les mêmes enzymes. La construction résultante a été électroporée dans *E. coli* DH5 $\alpha$  et *M. smegmatis* mc2 155. Un clone *M. smegmatis* émettant de la lumière et présentant  
 15 une activité *phoA* a été sélectionné et appelé pJVED/*blaF*. L'insert a été retiré en utilisant *Bam*HI et la construction refermée sur elle-même, reconstruisant ainsi le pJVEDa. Afin d'obtenir le pJVED<sub>b,c</sub>, le multisite de clonage a été coupé avec *Sca*I et *Kpn*I et refermé en enlevant un (pJVED<sub>b</sub>)  
 20 ou deux (pJVED<sub>c</sub>) nucléotides du site *Sna*BI. Après fusion, six cadres de lecture ont pu ainsi être obtenus. L'insert du pJVED/*hsp18* a été obtenu par amplification en chaîne par polymérase (ACP) de pPM1749 (Servant et al., 1995) en utilisant des oligonucléotides de la séquence :

- 25 18.Fw : 5'GTACCAGTATTGATCAGCTGTCTCCGAC3'  
 18.Back : AGTCAGGTACCTCGCGGAAAGGGTCAGTCCG3'

Le produit a été coupé avec *Kpn*I et *Sca*I, et ligaturé à pJVEDa, coupé avec les mêmes enzymes, donnant ainsi le pJVED/*hsp18*.

multisite de clonage de pJVED<sub>a</sub>.

#### Mesure de l'activité phosphatase alcaline

La présence d'activité est détectée par la couleur  
5 bleue des colonies croissant sur un milieu de culture  
contenant le substrat 5-bromo 4-chloro 3-indolyl phosphate  
(XP), puis l'activité peut être mesurée quantitativement de  
manière plus précise de la façon suivante :

*M. smegmatis* ont été cultivés dans un milieu LB  
10 additionnés de 0,05 % de Tween 80 (Aldrich) et de  
kanamycine (20 µg/ml<sup>-1</sup>) à 37°C pendant 24 heures.  
L'activité de la phosphatase alcaline a été mesurée par la  
méthode de Brockman et Heppel (Brockman et al., 1968) dans  
un extrait soniqué, avec p-nitrophénylphosphate comme  
15 substrat de la réaction. La quantité de protéines a été  
mesurée par essai Bio-Rad. L'activité phosphatase alcaline  
est exprimée en unité arbitraire (densité optique à 420 nm  
x µg de protéines<sup>-1</sup> x minutes<sup>-1</sup>).

#### 20 Mesure de l'activité luciférase

*M. smegmatis* a été cultivé dans un milieu LB  
additionné de 0,05 % de Tween 80 (Aldrich) et de kanamycine  
(20 µg/ml<sup>-1</sup>) à 37°C pendant 24 heures et utilisé en pleine  
25 croissance exponentielle (DO à 600 nm comprise entre 0,3 et  
0,8). Les aliquots de suspensions bactériennes ont été  
brièvement soniqués et l'extrait cellulaire a été utilisé  
pour mesurer l'activité de la luciférase. 25 µl de  
l'extrait soniqué ont été mélangés avec 100 µl de substrat  
30 (système d'essai luciférase Promega) automatiquement dans  
un luminomètre et la lumière émise exprimée en ULR ou RLU  
(Unités Lumineuses Relatives).

Construction de banques génomiques de *M. tuberculosis* et de *M. bovis*-BCG

5 Les banques ont été obtenues en utilisant essentiellement pJVEDa,b,c précédemment décrits.

Préparation de macrophages issus de la moelle osseuse et infection par *M. smegmatis* recombinants

10 Les macrophages issus de la moelle osseuse ont été préparés comme décrits par Lang et al., 1991. En résumé, les cellules de la moelle osseuse ont été prélevées du fémur de souris C57BL/6 âgée de 6 à 12 semaines (Iffa-Credo, France). Les cellules en suspensions ont été lavées et resuspendues dans du DMEM enrichi avec 10 % de sérum foetal de veau, 10 % de milieu L-cell conditionné et 2 mM de glutamine, sans antibiotiques. 10<sup>5</sup> cellules ont été ensemencées sur des plaques 24 puits Costar à fond plat dans 1 ml. Après quatre jours à 37°C dans une atmosphère humide à 10 % de teneur en CO<sub>2</sub>, les macrophages ont été rincés et réincubés pendant deux à quatre jours supplémentaires. Les cellules d'un puits contrôle ont été lysées avec du triton x 100 à 0,1 % dans l'eau et les noyaux énumérés. Environ 5 x 10<sup>5</sup> cellules adhérentes ont été comptées. Pour l'infection, *M. smegmatis* portant les différents plasmides a été cultivé en pleine phase exponentielle (DO<sub>600nm</sub> entre 0,4 et 0,8) et dilué jusqu'à une DO de 0,1 puis 10 fois dans un milieu pour macrophage. 1 ml a été ajouté à chaque puits et les plaques ont été centrifugées et incubées quatre heures à 37°C. Après trois lavages, les cellules

lyse (Promega). 100  $\mu$ l ont été soniqués et la lumière émise a été mesurée sur 25  $\mu$ m. Simultanément, les bactéries ont été énumérées par étalement sur L-agar-kanamycine (20  $\mu$ g/ml<sup>-1</sup>). La lumière émise est exprimée en ULR/10<sup>3</sup> bactéries.

#### Analyses des banques de données

Les séquences nucléotidiques ont été comparées à EMBL et GenBank en utilisant l'algorithme FASTA et les séquences protéiques ont été analysées par similitude grâce aux banques de données PIR et Swiss Prot en utilisant l'algorithme BLAST.

#### Exemple 1 : Les vecteurs pJVED

Les vecteurs pJVED (Figure 51) sont des plasmides portant un gène *phoA* tronqué de *E. coli* dépourvu de codon d'initiation, de séquence signal et de séquence régulatrice. Le site multiple de clonage (SMC) permet l'insertion de fragments des gènes codants pour d'éventuelles protéines exportées ainsi que leurs séquences de régulation. Dès lors, la protéine de fusion peut être produite et présenter une activité phosphatase alcaline si elle est exportée. Seules les fusions en phase pourront être productives. Ainsi, le SMC a été modifié de sorte que les fusions peuvent être obtenues dans six phases de lecture. En aval de *phoA*, le gène *luc* de la luciférase de luciole a été inséré. Le gène complet avec le codon d'initiation mais sans qu'aucun promoteur n'ait été utilisé devrait ainsi s'exprimer avec *phoA* comme dans un opéron synthétique. Un nouveau site de clonage des séquences

*coli-mycobacterium* avec un gène de résistance à la kanamycine comme marqueur de sélection.

*phoA* et *luc* fonctionnent comme dans un opéron, mais  
5 l'exportation est nécessaire pour l'activité *phoA*.

Quatre plasmides ont été construits par insertion dans le SMC de fragments d'ADN d'origine diverse :

Dans la première construction nommée pJVED/*blaF*, le fragment de 1,4 kb provient du plasmide déjà décrit pLA71  
10 (Lim et al., 1995). Ce fragment issu du gène  $\beta$ -lactamase (*blaF*) de *M. fortuitum* D216 (Timm et al., 1994) inclut le promoteur muté hyperactif, le segment codant pour 32 acides aminés de la séquence signal et les 5 premiers acides aminés de la protéine mature. Ainsi cette construction  
15 inclut le promoteur le plus fort connu chez *mycobacterium* et les éléments nécessaires à l'exportation de la protéine de la fusion *phoA*. Par conséquent, on peut attendre de cette construction une forte émission de lumière et une bonne activité *phoA* (cf figures 53 et 54).

20 Dans une deuxième construction nommée pJVED/*hsp18*, un fragment de 1,5 kb a été cloné à partir du plasmide déjà décrit pPM1745 (Servant et al., 1995). Ce fragment inclut les nucléotides codants pour les dix premiers acides aminés de la protéine de choc thermique de 18 kb issue de  
25 *Streptomyces albus* (heat shock protein 18, HSP 18), le site de liaison du ribosome, le promoteur et, en amont, des sites régulateurs contrôlant son expression. Cette protéine appartient à la famille de alpha crystalline de HSP à faible poids moléculaire (Verbon et al., 1992). Son  
30 homologue issu de *M. leprae*, l'antigène de 18 kDa, est déjà connu pour être actif.

On a pu constater que les deux constructions pJVED/*hsp18* et pJVED/*blaF* ont une activité *phoA* comparable à celle des constructions pJVED/*phoA* et pJVED/*phoA* + *phoA*.

Dans une troisième construction, nommée pJVED/P19kDa, l'insert issu de pExp410 (Lim et al., 1995) a été coupé et cloné dans le SMC de pJVEDa. Ce fragment inclut les nucléotides codants pour les 134 premiers acides aminés de la protéine connue de *M. tuberculosis* 19 kDa et de ses séquences régulatrices. Comme cela a pu être mis en évidence, cette protéine est une lipoprotéine glycosylée (Garbe et al., 1993 ; Herrmann et al., 1996). Sur les figures 53 et 54, on observe, pour cette construction, une bonne activité luc correspondant à un promoteur fort, mais l'activité *phoA* est la plus forte des quatre constructions. L'activité *phoA* élevée de cette protéine de fusion avec une lipoprotéine s'explique par le fait qu'elle reste attachée à la paroi cellulaire par son extrémité N-terminal.

Dans la quatrième et dernière construction nommée pJVED/erp l'insert provient de pExp53 (Lim et al., 1995) et a été cloné dans le SMC de pJVEDa. pExp53 est le plasmide initial sélectionné pour son activité *phoA* et contenant une partie du gène *erp* de *M. tuberculosis* qui code pour un antigène de 28 kDa. Ce dernier inclut la séquence signal, une partie de la protéine mature et, en amont du codon d'initiation, le site de liaison de ribosome. Le promoteur a été cartographié. Une boîte fer (iron box) putative du type fur est présente dans cette région et encadre la région -35 du promoteur (Berthot et al., 1995). Comme prévu (figures 53 et 54) cette construction présente une bonne émission lumineuse et une bonne activité *phoA*. Le fait que cette protéine de fusion, contrairement à la fusion avec la lipoprotéine de 19 kDa, ne semble pas attachée à la paroi cellulaire n'exclut pas que la protéine native...



*M. tuberculosis* dans les vecteurs pJVEDs et identification d'un des membres de ces banques, (DP428), induit au cours de la phagocytose par les macrophages murins dérivés de la moelle osseuse.

5 Les différentes constructions sont testées pour leur capacité à évaluer l'expression intracellulaire des gènes identifiés par l'expression de *phoA*. Dans cet objectif, l'activité *luc* est exprimée en URL pour 103 bactéries en culture axénique et/ou dans des conditions  
10 intracellulaires. L'induction ou la répression suivant la phagocytose par les macrophages murins dérivés de la moelle osseuse peut être évaluée convenablement par la mesure des activités spécifiques. Les résultats de deux expériences distinctes sont présentés dans le tableau 2.

15 Le plasmide pJVED/*hsp18* a été utilisé comme contrôle positif pour l'induction durant la phase de croissance intracellulaire. Bien que l'induction du promoteur par le chauffage de la bactérie à 42°C n'ait pas été concluant la phagocytose de la bactérie conduit  
20 clairement à une augmentation de l'activité du promoteur. Dans toutes les expériences, l'activité *luc* intracellulaire a été fortement induite, augmentant de 20 à 100 fois l'activité basale initialement faible (Servant, 1995).

Le plasmide pJVED/*blaF* a été utilisé comme contrôle  
25 de la modulation non spécifique au cours de la phagocytose. De faibles variations ont pu être mises en évidence, probablement dues à des changements de conditions de cultures. Quoi qu'il en soit, ces faibles variations ne sont pas comparables à l'induction observée avec le  
30 plasmide pJVED/*hsp18*.

Tous les membres de la banque d'ADN ont été analysés.

TABLEAU 1

Construction	% Récupération		URI/10 <sup>3</sup> bactéries extracellulaire	URI/10 <sup>3</sup> bactéries intracellulaire		Induction	
pJVED/ <i>blaI</i> * pJVED/ <i>hsp18</i> pJVED/ <i>DP428</i>	0,5 0,6 0,7		1460 8 0,06	1727 57 18		1,2 7,1 300	
Construction	% Récupération C57BL/6 Balb/C		URI/10 <sup>3</sup> bactéries extracellulaire	URI/10 <sup>3</sup> bactéries intracellulaire C57BL/6 Balb/C		Induction C57BL/6 Balb/C	
pJVED/ <i>blaI</i> * pJVED/ <i>hsp18</i> pJVED/ <i>DP428</i>	7 6,7 1,6	1,1 1,7 2,1	662 164 0,08	250 261 1,25	911 325 3,3	0,4 1,6 15,6	1,4 2 41

5 TABLEAU 2

Construction	% Récupération		URL/10 <sup>3</sup> bactéries extracellulaire	URL/10 <sup>3</sup> bactéries intracellulaire	Induction	
pJVED/ <i>blaI</i> *	22		1477	367	0,25	
pJVED/ <i>hsp18</i>	7		0,26	6,8	26	
pJVED/ <i>DP428</i>	21		0,14	4	28	

Le fragment nucléotidique codant pour la région N terminale du polypeptide DP428 de séquence SEQ ID N° 28 est contenu dans le plasmide déposé à la CEM sous le N° 1818.

La totalité de la séquence codant pour le polypeptide DP428 a été obtenue comme détaillée ci-après.

Une sonde a été obtenue par PCR à l'aide des oligonucléotides de séquence se

sonde. Les résultats de l'hybridation ont fait apparaître qu'un fragment d'ADN d'environ 1,7 kb était marqué. Du fait qu'il existe un site *Scal* s'étendant du nucléotide nt 984 au nucléotide nt 989 de la séquence SEQ ID N° 1, c'est-à-dire du côté 5' de la séquence utilisée comme sonde, la fin de la séquence codante est nécessairement présente dans le fragment détecté par hybridation.

L'ADN génomique de la souche *Mt* 103 de *M. tuberculosis*, après digestion par *Scal*, a subi une migration sur un gel d'agarose. Les fragments de tailles comprises entre 1,6 et 1,8 kb ont été clonés dans le vecteur pSL1180 (Pharmacia) préalablement clivé par *Scal* et déphosphorylé. Après transformation de *E. coli* avec les vecteurs recombinants résultants, les colonies obtenues ont été criblées à l'aide de la sonde. Le criblage a permis d'isoler six colonies hybridant avec cette sonde.

Les inserts contenus dans les plasmides des clones recombinants précédemment sélectionnés ont été séquencés, puis les séquences alignées de manière à déterminer la totalité de la séquence codant pour DP428, plus spécifiquement la SEQ ID N° 2.

Un couple d'amorces a été synthétisé afin d'amplifier, à partir de l'ADN génomique de *M. tuberculosis*, souche *Mt* 103, la totalité de la séquence codant pour le polypeptide DP428. L'amplicon obtenu a été cloné dans un vecteur d'expression.

Des couples d'amorces appropriés pour l'amplification et le clonage de la séquence codant pour le polypeptide DP428 peuvent être aisément réalisés par l'homme du métier, sur la base des séquences nucléotidiques SEQ ID N°1 et SEQ ID N°2.

Un couple d'amorces particulier peut être obtenu.

- Amorce aller (SEQ ID N° 29), comprenant la séquence allant du nucléotide en position nt 1021 au nucléotide nt 1044 de la séquence SEQ ID N° 2 :

5' -AGTGCATGCTGCTGGCCGAACCATCAGCGAC- 3'

5

- Amorce retour (SEQ ID N° 30), comprenant la séquence complémentaire de la séquence allant du nucléotide en position nt 1345 au nucléotide en position nt 1325 de la séquence SEQ ID N° 2 :

10 5' -CAGCCAGATCTGCGGGCGCCCTGCACCGCCTG- 3',

15 dans lesquelles la partie soulignée représente les séquences hybridant spécifiquement avec la séquence SEQ ID N° 2 et les extrémités 5' correspondent à des sites de restriction en vue du clonage de l'amplicon résultant dans un vecteur de clonage et/ou d'expression.

Un vecteur particulier utilisé pour l'expression du polypeptide DP428 est le vecteur pQE70 commercialisé par la société Qiagen.

20

**Exemple 3 :** La séquence complète du gène DP428 et de ses régions flanquantes.

25 Une sonde de la région codante de DP428 a été obtenue par ACP, et utilisée pour hybrider l'ADN génomique de différentes espèces de mycobactéries. D'après les résultats de la figure 3, le gène est présent uniquement dans les mycobactéries du complexe de *M. tuberculosis*.

30 L'analyse de la séquence suggère que DP428 pourrait faire partie d'un opéron. La séquence codante et les régions flanquantes ne présentent aucune homologie avec des séquences connues déposées dans les banques de données.

La séquence complète du gène DP428 est présentée dans l'annexe 1.

arginines pourraient correspondre à un signal de rétention et la protéine DP428 pourrait être accrochée par ce motif à des peptidoglycanes comme cela a déjà été décrit chez d'autres bactéries Gram<sup>+</sup> (Navarre et al., 1994 et 1996).

5 Le mécanisme de survie et de croissance intracellulaire des mycobactéries est complexe et les relations intimes entre la bactérie et la cellule hôte restent inexpliquées. Quel que soit le mécanisme, la croissance et la survie intracellulaire des mycobactéries  
10 dépend de facteurs produits par la bactérie et capables de moduler la réponse de l'hôte. Ces facteurs peuvent être des molécules exposées à la surface cellulaire telle que LAM ou des protéines associées à la surface cellulaire, ou des molécules activement secrétées.

15 D'un autre côté, intracellulairement, les bactéries elles-mêmes doivent faire face à un environnement hostile. Elles semblent y répondre par des moyens proches de ceux mis en oeuvre dans les conditions de stress, par l'induction de protéines de choc thermique (Dellagostin et  
20 al., 1995), mais aussi par induction ou la répression de différentes protéines (Lee et al., 1995). En utilisant une méthodologie dérivée de la PCR, Plum et Clark-curtiss (Plum et al., 1994) ont montré qu'un gène de *M. avium* inclu dans un fragment d'ADN de 3 kb, est induit après la phagocytose  
25 par des macrophages humains. Ce gène code pour une protéine exportée comprenant une séquence leader mais ne présentant pas d'homologie significative avec les séquences proposées par les banques de données. L'induction, pendant la phase de croissance intracellulaire, d'une protéine de choc  
30 thermique de faible poids moléculaire issue de *M. leprae* a également été mise en évidence (Dellagostin et al., 1995). Dans une autre étude, les protéines bactériennes de

induites et 28 réprimées. Les mêmes protéines sont mises en jeu au cours de stress provoqué par un faible pH, un choc thermique, H<sub>2</sub>O<sub>2</sub>, ou au cours de la phagocytose par des monocytes humains de la lignée THP1. Quoi qu'il en soit, le comportement des protéines induites et réprimées était unique dans chaque condition (Lee et al., 1995). Pris ensemble, ces résultats indiquent qu'un dialogue moléculaire subtile est mis en place entre les bactéries et leurs hôtes cellulaires. De ce dialogue dépend probablement le sort de l'organisme intracellulaire.

Dans ce contexte, l'induction de l'expression de DP428 pourrait être d'une importance majeure, indiquant un rôle important de cette protéine dans la survie et la croissance intracellulaire.

La méthode utilisée dans ces expériences pour évaluer l'expression intracellulaire des gènes (cf. Jacobs et al., 1993, pour la méthode de détermination de l'expression de la luciférase de luciole, et Lim et al., 1995, pour la méthode de détermination de l'expression du gène *PhoA*) présente l'avantage d'être simple comparée aux autres techniques comme la technique décrite par Mahan et al. (Mahan et al., 1993) adaptée aux mycobactéries et proposée par Bange et al. (Bange et al., 1996), ou la méthode subtractive basée sur l'ACP décrite par Plum et Clark curtiss (Plum et al., 1994). Il existe indiscutablement une variabilité comme le montre la comparaison des différentes expériences. Bien que provoquer l'induction ou la répression soit suffisant, il est désormais possible de l'évaluer fournissant ainsi un outil supplémentaire d'études physiologiques des protéines exportées identifiées par fusion avec *phoA*.

Des macrophages de moelle osseuse de souris sont préparés comme décrit par Lang et Antoine (Lang et al., 1991). Les bactéries de *M. segmentis* recombinantes, dont on a déterminé l'activité luciférase par  $10^3$  bactéries comme précédemment, sont incubées à 37°C sous atmosphère humidifiée et enrichie en CO<sub>2</sub> à 5%, pendant 4 heures en présence de ces macrophages de telle manière qu'elles soient phagocytées. Après rinçage pour éliminer les bactéries extracellulaires restantes, on ajoute au milieu de culture de l'amikacine (100 µg/ml) pendant deux heures. Après un nouveau rinçage, le milieu est remplacé par un milieu de culture (DMEM enrichi de 10 % de sérum de veau et 2 mM de glutamine) sans antibiotiques. Après une nuit d'incubation comme précédemment, les macrophages sont lysés à froid (4°C) à l'aide d'un tampon de lyse (cee lysis buffer, Promega), et l'activité luciférase par  $10^3$  bactéries déterminée. Le rapport des activités à la mise en culture et après une nuit donne le coefficient d'induction.

#### Exemple 5 :

Isolement d'une série de séquences par séquençage directement à partir des colonies.

Une série de séquences permettant l'expression et l'exportation de *phoA* ont été isolées à partir de l'ADN de *M. Tuberculosis* ou de *M. Bovis* BCG. Parmi ce groupe de séquences, deux d'entre elles ont été d'avantage étudiées, les gènes entiers correspondant aux inserts ont été clonés, séquencés, et des anticorps contre le produit de ces gènes ont servi à montrer en microscopie électronique que ces gènes codent pour une protéine exportée.

Une série de séquences a été isolée à partir de colonies de *M. Tuberculosis* exprimant une protéine exportée. Après la séquence, un

autre gène DP428 a été séquencé avant que la séquence du génome de *M. Tuberculosis* ne soit disponible. Il contient une séquence ressemblant à la séquence consensus d'attachement au peptidoglycane, ce qui suggère qu'il s'agit aussi d'un antigène vraisemblablement retrouvé à la surface des bacilles de la tuberculose. L'étude des trois gènes *erp*, *des*, et celui codant pour DP428 montre que le système *phoA* que nous avons développé chez les mycobactéries permet de repérer des gènes codant pour des protéine exportées sans déterminant repérable par des études *in silico*. Ceci est particulièrement vrai pour les polypeptides qui ne possèdent pas de séquence signal consensus (*des*) ou non pas de similarité avec des protéines de fonction connue (*erp* et DP428).

Un certain nombre d'inserts ont été identifiés et séquencés avant la connaissance du génome de *M. Tuberculosis*, d'autres après. Ces séquences peuvent être considérées comme des amorces permettant de rechercher des gènes codant pour des protéines exportées. A ce jour, une série d'amorces ont été séquencées et les gènes entiers correspondants ont été soit séquencés, soit identifiés d'après la séquence publiée du génome. Pour tenir compte des erreurs de séquençage toujours possibles, les régions en amont ou en aval de certaines amorces ont été considérées comme pouvant faire partie de séquences codant pour des protéines exportées. Dans certains cas des similarités avec des gènes codant pour des protéines exportées ou des séquences caractéristiques de signaux d'exportation ou des caractéristiques topologiques de protéines membranaires ont été détectées.

Les gènes identifiés par cette méthode peuvent être classés en deux catégories : ceux qui ont une similarité avec une amorce connue pour



des protéines exportées. C'est le cas de la séquence SEQ ID N° 8G et SEQ ID N° 8H possédant plus de 77 % de similarité avec SEQ ID N° 8A'.

Les séquences pouvant coder pour des protéines exportées sont les suivantes : SEQ ID N° 1, 8, 9, 8G, 8H, 13, 3, 10, 19, 20, 6, 16, 22, 23, 24, 39, 44, 46, et 50.

Des gènes identifiés d'après les amorces à partir de la séquence du génome n'ont aucune caractéristique (d'après la séquence) de protéines exportées. Il s'agit des séquences suivantes : SEQ ID N° 4, 27, 11, 12, 14, 7, 15, 17, 18, 21, 31, 32, 33, 34, 35, 36, 37, 38, 40, 41, 42, 43, 45, 47, 48, et 49.

D'après la séquence d'autres organismes comme *E. coli*, on peut rechercher dans la séquence du génome de *M. tuberculosis*, des gènes possédant des similarités avec des protéines connues pour être exportées chez d'autres organismes bien que ne possédant pas de séquence signal d'exportation. Dans ce cas une fusion avec *phoA* est un protocole avantageux pour déterminer si ces séquences de *M. tuberculosis* codent pour des protéines exportées bien que ne présentant pas de séquence signal consensus. Il a été en effet possible de cloner SEQ ID N° 49, une séquence similaire à un gène de *E. coli* de la famille *htrA*. Une fusion de SEQ ID N° 49 avec *phoA* conduit à l'expression et à l'exportation de *phoA*. Des colonies *M. smegmatis* hébergeant une fusion SEQ ID N° 49 *phoA* sur un plasmide pJVED sont bleues.

SEQ ID N° 49 est donc considérée comme une protéine exportée.

Les séquences suivantes ont été identifiées comme possédant des caractéristiques des protéines exportées.

Même si une séquence possède des déterminants de protéines exportées, cela ne démontre pas une exportation fonctionnelle. Le système *phoA* permet de montrer que le gène suspecté code réellement pour une protéine exportée.

- 5 Ainsi, il a été vérifié que la séquence SEQ ID N° 50 possédait bien des signaux d'exportation.

TABLEAU 3

SEQ ID N°	Référence de la séquence correspondante prédite par Cole et al.		Annotation
SEQ ID N°1	Rv 0203	.	Séquence hydrophobe en N-terminal
SEQ ID N°4			
SEQ ID N°27	Rv 2050		Pas de prédiction
SEQ ID N°8			
SEQ ID N°9	Rv 2563	.	Protéine membranaire
SEQ ID N°			
8G',H'	Rv 0072	.	Possible protéine de transport transmembranaire de type ABC
SEQ ID N°11	Rv 0546c	ML	Protein S-D Lactoyl Glutathione-méthyl glyoxal lyase
SEQ ID N°12	pas de prédiction		non retrouvé dans <i>M.tuberculosis</i> H37rv
SEQ ID N°13			
SEQ ID N°3	Rv 1384c	.	probable précurseur cutinase avec une séquence signal N-terminale
SEQ ID N°10			
SEQ ID N°14			
SEQ ID N°7	pas de prédiction		pas de prédiction
SEQ ID N°15	avec décalage de lecture, pourrait être en phase avec Rv 1530i		pas de prédiction
SEQ ID N°17	Rv 1303	ML	pas de prédiction
SEQ ID N°18	Rv 0199	ML	pas de prédiction
SEQ ID N°19	Rv 0419	.	site de fixation de lipide-protéine membranaire procaryote, similarité avec

			sérine/thréonine protéine kinase
SEQ ID N° 21	Rv 3365c	ML	similarité avec une métallo peptidase à zinc
SEQ ID N° 31	non prédite		pas de prédiction
SEQ ID N° 32	Rv 0822c	MI	Existence d'une région consensus avec la famille drac
SEQ ID N° 33	Rv 1044		pas de prédiction
SEQ ID N° 34	non prédite		pas de prédiction
SEQ ID N° 35	Rv 2169c		pas de prédiction
SEQ ID N° 36	Rv 3909	ML	pas de prédiction
SEQ ID N° 37	Rv 2753c		similarité avec des dihydropricolinate synthases
SEQ ID N° 38	Rv 0175		pas de prédiction
SEQ ID N° 39	Rv 3006	ML	prédiction de séquence signal de lipoprotéine
SEQ ID N° 40	Rv 0549c		pas de prédiction
SEQ ID N° 41	Rv 2975c pouvant être en phase avec Rv 2974c		similarité avec protéine de subtilis
SEQ ID N° 42	Rv 2622		similarité avec une méthyl transférase
SEQ ID N° 43	Rv 3278c	ML	pas de prédiction
SEQ ID N° 44	Rv 0309		pas de prédiction
SEQ ID N° 45	Rv 3193c	ML	pas de prédiction
SEQ ID N° 46	Rv 1411c		probable lipoprotéine avec une séquence signal N-terminale
SEQ ID N° 47	Rv 1714		similarité avec une glycine aldéhyde transase
SEQ ID N° 48	Rv 1331		similarité avec une protéine

SEQ ID N°5			
SEQ ID N°16	Rv 3810	• ML	Protéine de surface Berthelet et al. 1995
SEQ ID N°22 SEQ ID N°23 SEQ ID N°24	Rv 3763	•	Contient un site de fixation de lipoprotéine membranaire eucaryote
SEQ ID N°50	Rv 0125	•	Site actif des sérines protéases séquence signal N-terminale possible

Légende du tableau 3 :

Correspondance des séquences selon l'invention avec les  
séquences prédites par Cole et al. 1998, Nature, 393, 537-  
544.

\* : Prédiction que la protéine codée par la séquence  
soit exportée

ML : Prédiction de similarité avec *M. leprae*.

#### Exemple 6 :

Caractéristiques et obtention de la protéine M1C25

L'extrémité N terminale de la protéine M1C25 a été  
détectée par le système *PhoA* comme permettant l'exportation  
de la protéine de fusion, nécessaire à l'obtention de son  
activité phosphatase.

La séquence d'ADN codant pour l'extrémité N terminale de  
la protéine M1C25 est contenue dans la séquence SEQ ID N°  
20 de la présente demande de brevet.

A partir de cette séquence, on peut obtenir la protéine M1C25.

Rv3576,  
MTCY06G11.23,  
pknM

- 5 Séquence SEQ ID N° 29 du gène complet M1C25 (714 bases):  
cf. Figure 29

Ce gène code pour une protéine de 237 AA, de 25 kDa de  
masse molaire. Cette protéine est référencée dans les  
10 banques sous les appellations:

PID:e306716,  
SPTREMBL:P96858

- 15 Séquence SEQ ID N° 30 de la protéine M1C25 (237 acides  
aminés): cf. Figure 30

M1C25 contient un site de fixation à la partie  
lipidique des lipoprotéines de membrane des procaryotes  
(PS00013 Prokaryotic membrane lipoprotein lipid attachment  
20 site:

CTGGTCGGTG CGTGCATGCT CGCAGCCGGA TGC).

La fonction de M1C25 n'est pas certaine mais elle  
possède très probablement une activité "sérine/thréonine-  
protéine kinase". Des ressemblances sont à noter avec la  
25 moitié C terminale de K08G\_MYCTU Q11053 Rv1266c  
(MTCY50.16). Des similarités sont aussi retrouvées avec  
KY28\_MYCTU.

En 5' du gène codant pour M1C25 se trouve un gène  
codant potentiellement pour une protéine régulatrice  
30 (PID:e306715, SPTREMBL:P96857, Rv3575c, (MTCY06G11.22c))

Le profil d'hydrophobicité (Kyte et Doolittle) de M1C25  
est représenté à la figure 56.

Un site de clivage de la protéine

Le profil d'hydrophobicité (Kyte et Doolittle) de la protéine M1C25 est représenté à la figure 56.

d'hydrophobicité. Dans cette séquence signal potentielle il est a remarqué la répétition trois fois de la séquence des trois acides aminés LAA.

- 5 Clonage du gène M1C25 en vue de la production de la protéine qu'il code:

Un couple d'amorces a été synthétisé afin d'amplifier, à partir de l'ADN génomique de *M. tuberculosis*, souche  
10 H37Rv, la totalité de la séquence codant pour le polypeptide M1C25. L'amplicon obtenu a été cloné dans un vecteur d'expression.

Des couples d'amorces appropriés pour l'amplification  
15 et le clonage de la séquence codant pour M1C25 ont été synthétisés :

-amorce aller :

5' -ATAATACCATGGGCAAGCAGCTAGCCGCGC- 3'

-amorce retour :

20 5' -ATTTATAGATCTCTGCTTAGCAACCTTGGCCCGC- 3'

La partie soulignée représente les séquences hybridant spécifiquement avec la séquence M1C25 et les extrémités 5' correspondent à des sites de restriction en vue du clonage  
25 de l'amplicon résultant dans un vecteur de clonage et/ou d'expression.

Un vecteur particulier utilisé pour l'expression du polypeptide M1C25 est le vecteur pQE60 commercialisé par la société Qiagen, en suivant le protocole et les  
30 recommandations proposés par cette marque.

Les cellules utilisées pour le clonage sont des bactéries : *E. coli* XL1 Blue (résistante à la tétracycline).

La production de la protéine MC25 est illustrée par les figures 57 A et B. ( Extraits bactériens de la souche E. coli M15 contenant le plasmide pM1C25. Les cultures bactériennes et les extraits sont préparés selon Sambrook et al. (1989). L'analyse des extraits bactériens est effectuée selon les instructions de Quiagen (1997).

## Références bibliographiques

- AIDS therapies, 1993, in Mycobacterial infections, ISBN 0-9631698-1-5, pp. 1-11.
- 5 Altschul, S.F. et al., 1990, J. Mol. Biol., 215 : 403-410.
- Andersen, P. et al., 1991, Infect. Immun., 59 :1905-1910.
- Andersen, P. et al., 1995, J. Immunol., 154, 3359-3372.
- Bange, F.C., A.M. Brown, and W.R. Jacobs JR., 1996, Leucine auxotrophy restricts growth of *Mycobacterium bovis* BCG in
- 10 macrophages. Infect. Immun., 64,: 1794-1799.
- Barany, F., 1911, Proc. Natl. Acad. Sci. USA, 88 :189-193.
- Bates, J. et al., 1986, Am. Rev. Respir. Dis., 134 :415-417.
- Bates, J. 1979. Chest. 76(Suppl.):757-763.
- 15 Bates, J. et al., 1986. Am. Rev. Respir. Dis. 134 :415-417.
- Berthet, F.X., J. Rauzier, E.M. Lim, W. Philipp, B. Gicquel, and D. Portnoï, 1995, Characterization of the *M. tuberculosis* *crp* gene encoding a potential cell surface protein with repetitive structures. Microbiology. In press
- 20 Borremans, M. et al., 1989, Biochemistry, 7 : 3123-3130.
- Bouvet, E. 1994. Rev. Fr. Lab. 273 :53-56.
- Brockman, R.W. and Heppel L.A., 1968, On the localization of alkaline phosphatase and cyclic phosphodiesterase in *Escherichia coli*, Biochemistry, 7 : 2554-2561.
- 25 Burg, J.L. et al., 1995, Mol. and Cell. Probes, 10 :257-271.
- Chevrier, D. et al., 1993, Mol. and Cell. Probes, 7 :183-197.
- Clemens, D.



- Clemens, D.L. and Horwitz M.A., 1995, Characterization of the *Mycobacterium tuberculosis* phagosome and evidence that phagosomal maturation is inhibited, *J. Exp. Med.*, **181** : 257-270.
- 5 Colignon J.E., 1996. Immunologic studies in humans. Measurement of proliferative responses of cultured lymphocytes. *Current Protocols in Immunology*, NIH, 2. Section II.
- Daniel, T.M. et al. 1987. *Am. Rev. Respir. Dis.*, **135** :1137-1151).
- 10 Dellagostin, O.A., Esposito G., Eales L.-J., Dale J.W. and. McFadden J.J., 1995, Activity of mycobacterial promoters during intracellular and extracellular growth. *Microbiol.*, **141** : 2123-2130.
- 15 Drake, T.A. et al. 1987. *J. Clin. Microbiol.* **25**:1442-1445.
- Dramsfi et al., 1997, *Infection and Immunity*, **65**, 5 : 1615-1625.
- Duck, P. et al., 1990, *Biotechniques*, **9**:142-147.
- Erlich, H.A. 1989. In *PCR Technology. Principles and*
- 20 *Applications for DNA Amplification*. New York: Stockton Press.
- Felgner et al., 1987, *Proc. Natl. Acad. Sci.*, **84**:7413.
- Fraley et al., 1980, *J. Biol. Chem.*, **255**:10431.
- Gaillard, J.L., Berche P., Frohel C., Gouin E. and Cossart
- 25 P., 1991, Entry of *L. monocytogenes* into cells is mediated by internalin, a repeat protein reminiscent of surface antigens from Gram-positive cocci, *Cell*, **65** : 1127-1141.
- Garbe, T., Harris D., Vordermeier M., Lathigra R., Ivanyi J. and Young D., 1993, Expression of ...

- Guateli, J.C. et al., 1990, Proc. Natl. Acad. Sci. USA, 87:1874-1878.
- Harboe et al., 1996, Infect. Immun., 64, 16-22.
- Herrmann, J.L., O'Gaora P., Gallagher A., Thole J.E.R. and
- 5 Young D.B., 1996, Bacterial glycoproteins: a link between glycosylation and proteolytic cleavage of a 19 kDa antigen from *Mycobacterium tuberculosis*, EMBO J. 15 : 3547-3554.
- Houbenweyl, 1974, in Methode der Organischen Chemie, E. Wunsch Ed., Volume 15-I et 15-II, Thieme, Stuttgart.
- 10 Huygen, K. et al., 1996, Nature Medicine, 2(8):893-898.
- Innis, M.A. et al. 1990. in PCR Protocols. A guide to Methods and Applications. San Diego: Academic Press.
- Isberg, R.R., Voorhis D.L. and Falkow S., 1987, Identification of invasins: a protein that allows enteric .
- 15 bacteria to penetrate cultured mammalian cells, Cell, 50 : 769-778.
- Jacobs, W.R. et al., 1991. Construction of mycobacterial genomic libraries in shuttle cosmids. Genetic Systems for Mycobacteria, Methods in Enzymology, 204 : 537-555.
- 20 Jacobs, W.R. et al., 1993, Science, 260 :819-822.
- Kaneda, et al., 1989, Science, 243:375
- Kiehn, T.E., et al. 1987. J. Clin. Microbiol. 25 :1551-1552.
- Kievitis ,T. et al., 1991, J. Virol. Methods, 35 :273-286.
- 25 Kohler, G. et al., 1978, Nature, 256(5517):495-497.
- Kwoh, D.Y. et al., 1989, Proc. Natl. Acad. Sci. USA, 86 :1173-1177.
- Landegren ,U. et al., 1988, Science, 241,:1077-1080.
- Lang, T. and Antoine J.-C., 1998, .

- Lee, B.Y. and Horwitz M.A., 1995, Identification of macrophage and stress-induced proteins of *Mycobacterium tuberculosis*, J. Clin. Invest., 96 : 245-249.
- Lim, E.M., Rauzier J., Timm J., Torrea G., Murray A.,  
5 Gicquel B. and Portnoi D., 1995, Identification of *Mycobacterium tuberculosis* DNA sequences encoding exported proteins, using *phoA* gene fusions, J. Bacteriol., 177 : 59-65.
- Lizardi, P.M. et al., 1988, Bio/technology, 6 :1197-1202.
- 10 Mahan, M.J. et al., 1993. Selection of bacterial virulence genes that are specifically induced in host tissues, Science, 259 : 686-688.
- Manoil L., Mekalanos J.J. and Beckwith J., J. Bacteriol., 1990, 172, 515-518.
- 15 Matthews, J.A. et al., 1988, Anal. Biochem., 169:1-25.
- Merrifield, R.D., 1966, J. Am. Chem. Soc., 88(21):5051-5052.
- Midoux, 1993, Nucleic Acids Research, 21:871-878/
- Miole, E.A. et al., 1983, J. Mol. Biol., 171:281-295.
- 20 Minton, N.P., 1984, Gene, 31, 269-273.
- Montgomery et al., 1993, DNA Cell Biol., 12:777-783.
- Navarre, W.W. et al., 1994, Molecular Microbiology, 14(1):115-121.
- Navarre, W.W. et al., 1996, J. of Bacteriol., 178, 2 :441-446.
- 25 Pagano et al., 1967, J. Virol., 1 :891
- Pastore, 1994, Circulation, 90:1-517.
- Patel, et al. 1990, J. Clin. Microbiol., 28:1-519
- Patel, et al. 1990, J. Clin. Microbiol., 28:1-519

- Modulation of virulence factor expression by pathogen target cell contact. *Science.*, 273 : 1231-1233.
- Plum, G. and Clark-Curtiss J.E., 1994, Induction of *Mycobacterium avium* gene expression following phagocytosis by human macrophages. *Infect. Immun.*, 62 : 476-483.
- Roberts, M.C., et al. 1987. *J. Clin. Microbiol.* 25, :1239-1243.
- Rolfs, A. et al. 1991. In PCR Topics. Usage of Polymerase Chain reaction in Genetic and Infectious Disease. Berlin: Springer-Verlag.
- Sambrook, J. et al. 1989. In Molecular cloning : A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Sanchez-Pescador, R., 1988, *J. Clin. Microbiol.*, 26(10), :1934-1938.
- Schneewind, O. et al., 1995, *Science*, 268 : 103-106.
- Segev D., 1992, in « Non-radioactive Labeling and Detection of Biomolecules ». Kessler C. Springer Verlag, Berlin, New-York, 197-205.
- Servant, P. and Mazodier P., 1995, Characterization of *Streptomyces albus* 18-kilodalton heat shock-responsive protein. *J. Bacteriol.*, 177 : 2998-3003.
- Shiver, J.W., 1995, in Vaccines 1995, eds Chanock, R.M. Brown, F. Ginsberg, H.S. & Norrby, A., pp.95-98, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sorensen et al., 1995, *Infect. Immun.*, 63, 1710-1717.
- Stone, B.B. et al., 1995, *Mol. and Cell. Probes*, 10 : 355-370.
- Stover, C.E., Raneb, G.

- expressing outer surface protein A (OspA) lipoprotein: a candidate Lyme disease vaccine. *J. Exp. Med.*, 178 : 197-209.
- Sturgill-Koszycki, S., Schlesinger P.H., Chakroborty P., Haddix P.L., Collins H.L., Fok A.K., Allen R.D., Gluck S.L., Heuser J. and Russell D.G., 1994, Lack of acidification in *Mycobacterium* phagosomes by exclusion of the vesicular proton-ATPase. *Science.*, 263 : 678-681.
- Tascon, R.E et al., 1996, *Nature Medicine*, 2(8):888-892.
- 10 Technique assemblage oligonucléotides, 1983, *Proc. Natl. Acad. Sci. USA*, 80 :7461-7465.
- Technique des béta-cyanéthylphosphoramidites, 1986, *Bioorganic Chem.*, 4 :274-325.
- Thierry, D. et al., 1990, *Nucl. Acid Res.*, 18 :188.
- 15 Timm, J., Perilli M.G., Duez C., Trias J., Orefici G., Fattorini L., Amicosante G., Oratore A., Boris B., Frere J.M., Pugsley A.P. and Gicquel B., 1994, Transcription and expression analysis, using *lacZ* and *phoA* gene fusions, of *Mycobacterium fortuitum* B-lactamase genes cloned from a
- 20 natural isolate and a high-level B-lactamase producer. *Mol. Microbiol.*, 12 : 491-504.
- Tuberculosis Prevention Trial, 1981, Mendis, " Trial of BCG vaccines in South India for Tuberculosis Infection ", *Indian Journal of Medical research*, 1972 (Suppl.):1-74.
- 25 Urdea, M.S. et al., 1991, *Nucleic Acids Symp. Ser.*, 24 :197-200.
- Urdea, M.S., 1988, *Nucleic Acids Research*, 11: 4937-4957.
- Verbon, A., Hartskeerl R.A., Schuitema A., Kolk A.H., Young D.B. and Lathigra

- Walker, G.T. et al., 1992, Nucleic Acids Res., 20:1691-1696.
- Walker, G.T. et al., 1992, Proc. Natl. Acad. Sci. USA, 89:392-396.
- 5 Wiker, H.G. et al., 1992, Microbiol. Rev., 56 :648-661.
- Yamaguchi, R. et al., 1989, Infect. Immun., 57 :283-288 ;
- Xu, S., Cooper A., Sturgill-Koszycki S., van Heyningen T., Chatterjee D., Orme I., Allen P. and Russel D.G., 1994, Intracellular trafficking in *Mycobacterium tuberculosis* and
- 10 *Mycobacterium avium*-infected macrophages, J. Immuno., 153: 2568-2578.
- Young, D.B. et al., 1992, Mol. Microbiol., 6 :133-145.
- Yuen, L.K.W. et al., 1993, J. Clin. Microbiol., 31 : 1615-1618.

REVENDICATIONS

1. Vecteur recombinant de criblage, de clonage et/ou d'expression, caractérisé en ce qu'il se réplique chez des mycobactéries et en ce qu'il contient :
  - 1) un réplicon fonctionnel chez les mycobactéries ;
  - 2) un marqueur de sélection ;
  - 3) une cassette rapporteur comprenant :
    - a) un site de clonage multiple (polylinker),
    - b) éventuellement un terminateur de transcription actif chez les mycobactéries, en amont du polylinker,
    - c) une séquence nucléotidique codante issue d'un gène codant pour un marqueur d'expression, d'exportation et/ou de sécrétion de protéine, ladite séquence nucléotidique étant dépourvue de son codon d'initiation et de ses séquences de régulation, et
    - d) une séquence nucléotidique codante issue d'un gène codant pour un marqueur d'activité de promoteurs contenus dans le même fragment, ladite séquence nucléotidique étant pourvue de son codon d'initiation
2. Vecteur recombinant selon la revendication 1, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'expression, d'exportation et/ou de sécrétion de protéine est une séquence codante issue du gène *pH A* de la phosphatase alcaline.
3. Vecteur recombinant selon l'une des revendications 1 et 2, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'expression, d'exportation et/ou de sécrétion de protéine est une séquence codante issue du gène *pH A* de la phosphatase alcaline.

4. Vecteur recombinant selon l'une des revendications 1 à 3, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'activité de promoteurs contenus dans le même fragment est une séquence  
5 codante issue du gène *luc* de la luciférase de luciole.
5. Vecteur recombinant selon l'une des revendications 1 à 4, caractérisé en ce que la séquence nucléotidique codante issue d'un gène codant pour un marqueur d'activité de  
10 promoteurs contenus dans le même fragment est une séquence codante issue du gène *GFP* de la Green Fluorescent Protein.
6. Vecteur recombinant selon l'une des revendications 1 à 5, caractérisé en ce que le terminateur de transcription  
15 actif chez les mycobactéries est le terminateur du coliphage T4 (*tT4*).
7. Vecteur recombinant selon l'une des revendications 1 à 6, caractérisé en ce qu'il est un plasmide choisi parmi les  
20 plasmides suivants déposés à la CNCM (Collection Nationale de Cultures de Microorganismes, Paris, France) :
- a) pJVEDa déposé à la CNCM sous le N° I-1797, le 12/12/1996,
- b) pJVEDb déposé à la CNCM sous le N° I-1906, le 25 juillet  
25 1997,
- c) pJVEDc déposé à la CNCM sous le N° I-1799, le 12/12/1996.
8. Vecteur recombinant selon l'une des revendications 1 à  
30 7, caractérisé en ce qu'il comprend en l'un des sites de clonage du polylinker une séquence d'acide nucléique de mycobactérie chez laquelle on détecte un site de susceptibilité.



favoriser l'exportation et/ou la sécrétion dudit polypeptide, ou tout ou partie de gène codant pour ledit polypeptide.

- 5 9. Vecteur recombinant selon l'une des revendications 1 à 8, caractérisé en ce que la séquence d'acide nucléique de mycobactérie qu'il contient est obtenue par fragmentation physique ou par digestion enzymatique de l'ADN génomique ou de l'ADN complémentaire d'un ARN d'une mycobactérie.
- 10 10. Vecteur recombinant selon l'une des revendications 1 à 9, caractérisé en ce que ladite mycobactérie est *M. tuberculosis*.
- 15 11. Vecteur recombinant selon l'une des revendications 1 à 9, caractérisé en ce que ladite mycobactérie est choisie parmi *M. africanum*, *M. bovis*, *M. avium* ou *M. leprae*.
- 20 12. Vecteur recombinant selon la revendication 10, caractérisé en ce qu'il est un plasmide choisi parmi les plasmides suivants déposés à la CNCM :
- a) p6D7 déposé le 28 janvier 1997 à la CNCM sous le N°I-1814,
- b) p5A3 déposé le 28 janvier 1997 à la CNCM sous le N°I-1815,
- 25 c) p5F6 déposé le 28 janvier 1997 à la CNCM sous le N°I-1816,
- d) p2A29 déposé le 28 janvier 1997 à la CNCM sous le N°I-1817,
- 30 e) pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818,
- f) p5B5 déposé le 28 janvier 1997 à la CNCM sous le N°I-1819.

- i) p1B7 déposé le 31 janvier 1997 à la CNCM sous le N°I-1843,  
j) pJVED/*M. tuberculosis* déposé le 25 juillet 1997 à la CNCM sous le N° I-1907,  
5 k) pM1C25 déposé le 4 août 1998 à la CNCM sous le n°I-2062.

13. Vecteur recombinant selon la revendication 12, caractérisé en ce qu'il s'agit du plasmide pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818.

10

14. Procédé de criblage de séquences de nucléotides issues de mycobactéries pour déterminer la présence de séquences correspondant à des polypeptides exportés et/ou sécrétés pouvant être induits ou réprimés lors de l'infection, leurs  
15 séquences promotrices et/ou régulatrices associées susceptibles notamment de permettre ou de favoriser l'exportation et/ou la sécrétion desdits polypeptides d'intérêt, ou tout ou partie de gènes d'intérêt codant pour lesdits polypeptides, caractérisé en ce qu'il met en oeuvre  
20 un vecteur selon l'une des revendications 1 à 13.

15. Procédé de criblage selon la revendication 14, caractérisé en ce qu'il comprend les étapes suivantes :  
a) la fragmentation physique des séquences d'ADN de  
25 mycobactéries ou leur digestion par au moins une enzyme déterminée et la récupération des fragments obtenus ;  
b) l'insertion des fragments obtenus à l'étape a) dans un site de clonage, compatible le cas échéant avec l'enzyme de l'étape a), du polylinker d'un vecteur selon l'une des  
30 revendications 1 à 13 ;  
c) si besoin, l'amplification desdits fragments contenus dans le vecteur, par exemple par répllication de ce dernier après insertion du vecteur dans une cellule.

le vecteur est un vecteur d'expression.

- e) la culture de cellules hôtes transformées dans un milieu permettant la mise en évidence du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs contenu dans le vecteur ;
- 5 f) la détection des cellules hôtes positives (colonies positives) pour l'expression du marqueur d'exportation et/ou de sécrétion, et /ou du marqueur d'activité de promoteurs ;
- g) l'isolement de l'ADN des colonies positives et  
10 l'insertion de cet ADN dans une cellule identique à celle de l'étape c) ;
- h) la sélection des insertions contenues dans le vecteur, permettant l'obtention de clones positifs pour le marqueur d'exportation et/ou de sécrétion, et /ou pour le marqueur  
15 d'activité de promoteurs ;
- i) l'isolement et la caractérisation des fragments d'ADN de mycobactéries contenues dans ces insérats, et l'étape i) du procédé pouvant comporter en outre une étape de séquençage des insertions sélectionnées.
- 20
16. Banque d'ADN génomique ou d'ARN<sup>m</sup> complémentaire d'ARN<sup>m</sup> de mycobactérie, caractérisée en ce qu'elle est obtenue par un procédé selon la revendication 14 et/ou un procédé comprenant les étapes a) et b), ou a), b) et c) du procédé  
25 selon la revendication 15.
17. Banque d'ADN génomique ou d'ARN<sup>m</sup> complémentaire d'ARN<sup>m</sup> de mycobactérie selon la revendication 16 caractérisée en ce que ladite mycobactérie est une mycobactérie pathogène.
- 30
18. Banque d'ADN génomique ou d'ARN<sup>m</sup> complémentaire d'ARN<sup>m</sup> de mycobactérie selon la revendication 17, caractérisée en ce que ladite mycobactérie est

19. Banque d'ADN génomique ou d'ADNc complémentaire d'ARNm de mycobactérie selon la revendication 18, caractérisée en ce que ladite mycobactérie est *Mycobacterium tuberculosis*.
- 5 20. Séquence nucléotidique de mycobactérie ou comprenant une séquence nucléotidique de mycobactérie susceptible d'être sélectionnée par un procédé selon l'une des revendications 14 et 15.
- 10 21. Séquence nucléotidique de mycobactérie ou comprenant une séquence nucléotidique de mycobactérie selon la revendication 20, caractérisée en ce que ladite mycobactérie est choisie parmi *M. tuberculosis*, *M. bovis*, *M. africanum*, *M. avium*, *M. leprae*, *M. paratuberculosis*, *M.*  
 15 *kansassi* ou *M. xénopi*.
22. Séquence nucléotidique selon l'une des revendications 20-21 caractérisée en ce qu'elle est choisie parmi les séquences de fragments d'ADN de mycobactérie de séquence  
 20 nucléique SEQ ID N°1 à SEQ ID N° 24C, SEQ ID N°27A à SEQ ID N°27C, SEQ ID N°29 et SEQ ID N° 31A à SEQ ID N°50F.
23. Séquence nucléotidique de mycobactérie l'une des revendications 20-21 caractérisée en ce qu'elle est choisie  
 25 parmi les séquences de fragments d'ADN de mycobactérie de séquence SEQ ID N°1, SEQ ID N°3A, SEQ ID N°5A, SEQ ID N°6A, SEQ ID N°7A, SEQ ID N°8A, SEQ ID N°9A, SEQ ID N°10A, SEQ ID N°27A ou SEQ ID N°28 contenus respectivement dans les vecteurs pDP428 (CNCM, N°I-1818), pSD7 (CNCM, N°I-1814),  
 30 pSF6 (CNCM, N°I-1816), p2A29 (CNCM, N°I-1817), p5B5 (CNCM, N°I-1819), p1C7 (CNCM, N°I-1820), p2H7 (CNCM, N°I-1821), p1B7 (CNCM, N°I-1843), p5A3 (CNCM, N°I-1815) ou pMIC28.

## III

25. Polynucléotide caractérisé en ce qu'il comprend un polynucléotide choisi parmi :

- 5 a) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide selon l'une des revendications 20 à 24,
- b) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide selon l'une des revendications 20 à 24,
- 10 c) un polynucléotide hybridant dans des conditions de forte stringence avec une séquence de polynucléotide selon l'une des revendications 20 à 24,
- d) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide défini selon l'une des revendications 20 à 24 ou défini en a).
- 15

26. Polypeptide, leurs fragments ou fragments biologiquement actifs ou leurs polypeptides homologues, susceptible d'être codé par une séquence nucléotidique de mycobactérie selon l'une des revendications 20 à 25, et susceptible d'être exporté et/ou sécrété, et/ou induit ou réprimé, ou exprimé de façon constitutive lors de l'infection.

25 27. Mycobactérie recombinante caractérisée en ce qu'elle est transformée par un vecteur recombinant selon l'une des revendications 1 à 13.

28. Polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques de séquence SEQ ID N°1 à SEQ ID N°2.

29. Polynucléotide dont la séquence est choisie parmi les séquences nucléotidiques de séquence SEQ ID N°1 à SEQ ID N°2.

- b) un polynucléotide dont la séquence nucléique est la séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses, de la séquence SEQ ID N°1,
- 5 c) un polynucléotide dont la séquence est complémentaire de la séquence d'un polynucléotide défini en a) ou b),
- d) un polynucléotide dont la séquence comporte au moins 50% d'identité avec un polynucléotide défini en a), b) ou c),
- e) un polynucléotide hybridant dans des conditions de forte
- 10 stringence avec une séquence de polynucléotide défini en a), b), c) ou d),
- f) un fragment d'au moins 8 nucléotides consécutifs d'un polynucléotide défini en a), b), c), d) ou e).
- 15 30. Polynucléotide selon l'une des revendications 20 à 25, 28 et 29, caractérisé en ce que sa séquence nucléique hybride avec l'ADN de séquence de mycobactéries et préférentiellement avec de l'ADN de séquences de mycobactéries appartenant au complexe de *Mycobacterium*
- 20 *tuberculosis*.
31. Polypeptide caractérisé en ce qu'il est codé par une séquence polynucléotidique selon l'une des revendications 20 à 25.
- 25 32. Polypeptide caractérisé en ce qu'il comprend un polypeptide choisi parmi :
- a) un polypeptide dont la séquence d'acides aminés est comprise dans une séquence d'acides aminés choisie parmi
- 30 les séquences d'acides aminés SEQ ID N° 1 à SEQ ID N° 24C, SEQ ID N° 27A à SEQ ID N° 28 ou SEQ ID N° 30 à SEQ ID N° 50F,

33. Polypeptide dont la séquence d'acides aminés est comprise dans la séquence d'acides aminés SEQ ID N°1 ou SEQ ID N°2, ou polypeptide de séquence d'acides aminés SEQ ID N°28.
34. Polynucléotide caractérisé en ce qu'il code pour un polypeptide selon l'un des revendications 32 et 33.
35. Séquence d'acide nucléique utilisable comme amorce, caractérisée en ce que ladite séquence est choisie parmi les séquences d'acide nucléique de polynucléotide selon l'une des revendications 20 à 25, 28 à 30, et 34.
36. Séquence d'acide nucléique selon la revendication 35, caractérisée en ce que ladite séquence est choisie parmi les séquences SEQ ID N°25 et SEQ ID N°26.
37. Séquence d'acide nucléique selon l'une des revendications 35 et 36 pour la détection et/ou l'amplification de séquences nucléiques.
38. Séquence d'acide nucléique utilisable comme sonde, caractérisée en ce que ladite séquence est choisie parmi les séquences d'acide nucléique selon l'une des revendications 20 à 25, 28 à 30, et 34.
39. Séquence d'acide nucléique selon la revendication 38, caractérisée en ce qu'elle est marquée par un composé radioactif ou par un composé non radioactif.
40. Séquence d'acide nucléique selon l'une des

41. Séquence d'acide nucléique selon l'une des revendications 38 à 40 pour la détection et/ou l'amplification de séquences nucléiques.
- 5 42. Séquence d'acide nucléique selon l'une des revendications 38 à 41, caractérisée en ce que ladite séquence est une séquence comprise entre le nucléotide en position nt 964 et le nucléotide en position nt 1234, extrémités incluses, de la séquence SEQ ID N°1.
- 10 43. Vecteur recombinant de clonage, d'expression et/ou d'insertion, caractérisé en ce qu'il contient une séquence d'acide nucléique de polynucléotide selon l'une des revendications 20 à 25, 28 à 30, et 34.
- 15 44. Cellule hôte, caractérisée en ce qu'elle est transformée par un vecteur recombinant selon la revendication 43.
- 20 45. Cellule hôte selon la revendication 44, caractérisée en ce qu'il s'agit de la souche de *E. coli* transformée par le plasmide pDP428 déposé le 28 janvier 1997 à la CNCM sous le N°I-1818 ou transformée par le plasmide pMIC25 déposé le 4 août 1998 à la CNCM sous le n°I-2062, ou d'une souche de
- 25 *M. tuberculosis*, *M. bovis* ou *M. africanum* possédant potentiellement tous les systèmes de régulation appropriés.
46. Procédé de préparation d'un polypeptide, caractérisé en ce qu'il met en oeuvre un vecteur selon la revendication
- 30 43.
47. Polypeptide recombinant susceptible d'être obtenu par
48. Procédé de préparation d'un polypeptide, caractérisé en ce qu'il met en oeuvre un vecteur selon la revendication 43 et une séquence selon



polypeptide susceptible d'induire une réponse immunitaire chez l'homme ou l'animal.

49. Polypeptide hybride selon la revendication 48,  
5 caractérisé en ce que le polypeptide susceptible d'induire une réponse immunitaire contient au moins un déterminant antigénique capable d'induire une réponse humorale et/ou cellulaire.
- 10 50. Polynucléotide codant pour un polypeptide hybride selon l'une des revendications 48 et 49.
51. Polypeptide hybride selon l'une des revendications 48 et 49, caractérisée en ce qu'il s'agit d'une protéine  
15 recombinante obtenue par l'expression d'un polynucléotide selon la revendication 50.
52. Procédé pour la détection *in vitro* d'anticorps dirigés contre une mycobactérie et préférentiellement une bactérie  
20 du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :
- a) mise en contact de l'échantillon biologique avec un polypeptide selon l'une des revendications 26, 32, 33 et  
25 47 ;
- b) mise en évidence du complexe antigène-anticorps formé.
53. Procédé pour la détection d'une infection par une mycobactérie et préférentiellement une bactérie du complexe  
30 *Mycobacterium tuberculosis* dans un mammifère, caractérisé en ce qu'il comprend les étapes suivantes :
- a) préparation d'un échantillon biologique

- b) incubation de l'échantillon biologique de l'étape a) avec un polypeptide selon l'une quelconque des revendications 26, 32, 33 et 47 ;
- c) détection d'une réaction cellulaire indiquant une sensibilisation préalable du mammifère audit polypeptide notamment la prolifération cellulaire et/ou la synthèse de protéines telles que l'interféron gamma ;
- d) détection d'une réaction d'hypersensibilité retardée ou de sensibilisation du mammifère audit polypeptide.
54. Kit pour le diagnostic in vitro d'une infection par une mycobactérie appartenant au complexe *Mycobacterium tuberculosis*, comprenant :
- a) un polypeptide selon l'une des revendications 26, 32, 33 et 47 ;
- b) le cas échéant, les réactifs pour la constitution du milieu propice à la réaction immunologique ;
- c) les réactifs permettant la détection des complexes antigène-anticorps produits par la réaction immunologique ;
- d) le cas échéant, un échantillon biologique de référence (témoin négatif) dépourvu d'anticorps reconnus par ledit polypeptide ;
- e) le cas échéant, un échantillon biologique de référence (témoin positif) contenant une quantité prédéterminée d'anticorps reconnus par ledit polypeptide.
55. Anticorps mono- ou polyclonaux, leurs fragments, ou anticorps chimériques, caractérisés en ce qu'ils sont capables de reconnaître spécifiquement un polypeptide selon l'une des revendications 26, 32, 33 et 47 .
56. Anticorps selon la revendication 55, caractérisés en ce

tuberculosis dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

- a) mise en contact de l'échantillon biologique avec un anticorps selon l'une des revendications 55 et 56 ;
- 5 b) mise en évidence du complexe antigène-anticorps formé.

58. Kit pour la détection spécifique de la présence d'un antigène d'une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en

10 ce qu'il comprend les éléments suivants :

- a) un anticorps polyclonal ou monoclonal selon l'une des revendications 55 et 56 ;
- b) les réactifs pour la constitution du milieu propice à la réaction immunologique ;
- 15 c) les réactifs permettant la détection des complexes antigène-anticorps produits par la réaction immunologique.

59. Procédé de détection et d'identification rapide d'une mycobactérie et préférentiellement de *M. tuberculosis* dans

20 un échantillon biologique, caractérisé en ce qu'il comporte les étapes suivantes :

- a) isolement de l'ADN à partir de l'échantillon biologique à analyser, ou obtention d'un ADNc à partir de l'ARN de l'échantillon biologique ;
- 25 b) amplification spécifique de l'ADN des mycobactéries appartenant au complexe *Mycobacterium tuberculosis* à l'aide d'amorces selon l'une des revendications 35 à 37 ;
- c) analyse des produits d'amplification.

30 60. Procédé pour la détection de bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon

61. Procédé pour la détection de bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :

ayant, le cas échéant, préalablement été rendu accessible à l'hybridation, dans des conditions permettant l'hybridation de la sonde à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;

- 5 b) détection de l'hybride formé entre la sonde oligonucléotidique et l'ADN de l'échantillon biologique.

61. Procédé pour la détection de bactéries appartenant au complexe *Mycobacterium tuberculosis* dans un échantillon  
10 biologique, caractérisé en ce qu'il comprend les étapes suivantes :

a) mise en contact d'une sonde oligonucléotidique immobilisée sur un support selon la revendication 40 avec  
15 un échantillon biologique, l'ADN de l'échantillon, ayant, le cas échéant, été préalablement rendu accessible à l'hybridation, dans des conditions permettant l'hybridation de la sonde à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;

b) mise en contact de l'hybride formé entre la sonde  
20 oligonucléotidique immobilisée sur un support et l'ADN contenu dans l'échantillon biologique, le cas échéant après élimination de l'ADN de l'échantillon biologique n'ayant pas hybridé avec la sonde, avec une sonde oligonucléotidique marquée selon la revendication 32.

25 62. Procédé de détection selon la revendication 61, caractérisé en ce que, préalablement à l'étape a), l'ADN de l'échantillon biologique, ou l'ADNc obtenu par transcription inverse de l'ARN de l'échantillon, est  
30 amplifié à l'aide d'un couple d'amorces selon l'une des revendications 35 à 37.

- a) mise en contact de l'échantillon biologique avec un couple d'amorces selon l'une des revendications 35 à 37, l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des conditions permettant une hybridation des amorces à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;
- b) amplification de l'ADN de la bactérie du complexe *Mycobacterium tuberculosis* ;
- c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par les amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique marquée selon la revendication 39.
64. Procédé pour la détection de la présence d'une bactérie appartenant au complexe de *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les étapes suivantes :
- a) mise en contact de l'échantillon biologique avec deux couples d'amorces selon l'une des revendications 35 à 37, l'ADN contenu dans l'échantillon ayant été, le cas échéant, préalablement rendu accessible à l'hybridation, dans des conditions permettant une hybridation des amorces à l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* ;
- b) amplification de l'ADN de la bactérie du complexe *Mycobacterium tuberculosis* ;
- c) mise en évidence de l'amplification de fragments d'ADN correspondant au fragment encadré par lesdites amorces, par exemple par électrophorèse sur gel ou au moyen d'une sonde oligonucléotidique marquée selon la revendication 39.

- a) une sonde oligonucléotidique selon l'une des revendications 38 à 42 ;
- b) les réactifs nécessaires à la mise en oeuvre d'une réaction d'hybridation ;
- 5 c) le cas échéant, un couple d'amorces selon l'une des revendications 35 à 37 ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN (ADN génomique, plasmidique ou ADNc) d'une bactérie du complexe *Mycobacterium tuberculosis*.
- 10 66. Kit ou nécessaire pour la détection de la présence d'une bactérie du complexe *Mycobacterium tuberculosis* dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :
- 15 a) une sonde oligonucléotidique, dite sonde de capture, selon la revendication 40 ;
- b) une sonde oligonucléotidique, dite sonde de révélation, selon l'une des revendications 38 à 42 ;
- 20 c) le cas échéant, un couple d'amorces selon l'une des revendications 35 à 37 ainsi que les réactifs nécessaires à une réaction d'amplification de l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis*.
- 25 67. Kit pour l'amplification de l'ADN d'une bactérie du complexe *Mycobacterium tuberculosis* présent dans un échantillon biologique, caractérisé en ce qu'il comprend les éléments suivants :
- a) un couple d'amorces selon l'une des revendications 35 à 37 ;
- 30 b) les réactifs nécessaires pour effectuer une réaction d'amplification d'ADN ;
- c) éventuellement un composé chimique ou un produit biologique.

68. Composition immunogène caractérisée en ce qu'elle comprend un ou plusieurs polypeptides selon l'une des revendications 26, 32, 33 et 47 et/ou un ou plusieurs polypeptides hybrides selon l'une des revendications 48, 49  
5 et 51.

69. Vaccin caractérisé en ce qu'il contient un ou plusieurs polypeptides selon l'une des revendications 26, 32, 33 et 47 et/ou un ou plusieurs polypeptides hybrides selon l'une  
10 des revendications 48, 49 et 51, en association avec un véhicule pharmaceutiquement compatible et, le cas échéant un ou plusieurs adjuvants de l'immunité appropriés.

70. Vaccin destiné à l'immunisation à l'encontre d'une  
15 infection bactérienne ou virale, telle que la tuberculose ou l'hépatite, comprenant un vecteur selon la revendication 43 ou un polynucléotide selon la revendication 50, en association avec un véhicule pharmaceutiquement acceptable.

20 71. Vaccin caractérisé en ce qu'il contient une ou plusieurs séquences nucléotidiques selon l'une des revendications 20 à 24 et/ou un ou plusieurs polynucléotides selon la revendication 25 en association avec un véhicule pharmaceutiquement compatible et, le cas  
25 échéant, un ou plusieurs adjuvants de l'immunité appropriés.

72. Méthode de criblage de molécules capables d'inhiber la croissance de mycobactéries ou le maintien de  
30 mycobactéries dans un hôte, caractérisée en ce que lesdites molécules bloquent la synthèse ou la fonction des

73. Méthode de criblage selon la revendication 72, caractérisée en ce que les molécules sont des anti-messagers ou induisent la synthèse d'anti-messagers.

5        74. Molécules capables d'inhiber la croissance de mycobactéries ou le maintien de mycobactéries dans un hôte, caractérisées en ce que lesdites molécules sont synthétisées d'après la structure des polypeptides codés par une séquence nucléotidique selon l'une quelconque des  
10 revendications 20 à 24 ou par un polynucléotide selon la revendication 25.

15

20





2/185

Insert du clone contenant DP428 et contenu dans seq1  
1/1 31/11  
GAT CGC CTT TGA CGC CTA TTC GGT CGC GCA GCT TTT TGG CGA CGT CAC TGG TGC CCG CGT  
asp arg leu OPA arg leu phe gly arg ala ala phe trp arg arg his trp cys pro arg  
61/21 91/31  
TGC GGG CGT CCA GCC GCA GCG ACA CCA CAT ACG GCC GGT CCG GTT GCG GGG GCC GTT GGG  
cys gly arg pro ala ala ala thr pro his thr ala gly pro val ala gly ala val gly  
121/41 151/51  
TGG GGT TGG GTG CCT CCG TCA CCC CAG GCA GTT CGC TGG CTA TTT GTC GCA GTA GCG CGA  
trp gly trp val pro pro ser pro gln ala val arg trp leu phe val ala val ala arg  
181/61 211/71  
CGG CAT TGT CGA TGT CTT GGT AGC TAG CAT CCG GTC GGG GGG CCG CTA CCA GCG CCA GCG  
arg his cys arg cys leu gly ser AMB his pro val gly gly pro leu pro ala pro ala  
241/81 271/91  
CCG GGG CTC CCC GGT CCG GGT AGT GCG CGT CGA GTT GGT CGT GGA CCA GCA ATG ACT GCG  
pro gly leu pro gly pro gly ser ala arg arg val gly arg gly pro ala met thr ala  
301/101 331/111  
ACC GCG CGA CTT CGA AAC CGC CAC CGG TTA GAT TCC CCG ACT GCG TCA TCG CCA GGT AAA  
thr arg arg leu arg asn arg his arg leu asp ser pro thr ala ser ser pro gly lys  
361/121 391/131  
CCG CCG GCA CTA ACG CCA GCA ACC AAC CCG TCA AGA CCA ACC AAC GGC ACC TGC GCA GGT  
pro pro ala leu thr pro ala thr asn pro OPA arg pro thr asn gly thr cys ala gly  
421/141 451/151  
TGC GGC TCA ACC GCA TCA TGA ACT GCT GGA TTT CCG ACT CCC CGT ACT CTC GCG CAG TGT  
cys gly ser thr ala ser OPA thr ala gly phe arg thr pro arg thr leu ala gln cys  
481/161 511/171  
GTG CCC GCG AGC CTA CCG AAG ATC GCG TGC ATG CGT TCG GCG TGG ACC GCA CAG CAC CTG  
val pro ala ser leu pro lys ile ala cys met arg ser ala trp thr ala gln his leu  
541/181 571/191  
GAG TTG GCG GCG CCG AGG GCG GAG ATG GCA GGA TGA CCG ATT CTC GGG GGC GGG AAC TCG  
glu leu ala ala pro arg ala glu met ala gly OPA arg ile val gly gly gly asn ser  
601/201 631/211  
CAG GCG CCC GGA CCG TCG CAA ACC CGT CCG AAA CCC GTC GCA AAC CGT AAG GAG TCA TCC  
gln ala ala gly pro ser gln thr arg arg lys pro val ala asn arg lys glu ser ser  
661/221 691/231  
ATG AAG ACA GCG ACC GCG ACG ACG CCG CCG AGG CTC TTG GCA GTA CTG ATC GCG CTC GCG  
met lys thr gly thr ala thr thr arg arg arg leu leu ala val leu ile ala leu ala  
721/241 751/251  
TTG CCG GCG GCG GCG GTT GCG CTG CTG GCG GAA CCA TCA ACC AAC GCG GCG TCG GAC CCG  
leu pro gly ala ala val ala leu leu ala glu pro ser ala thr gly ala ser asp pro  
781/261 811/271  
TGC GCG GCG ACC GAA GTG GCG AGG ACC GTG GGT TCG GTT GCG AAC TCG ATG GGT GAT TA  
cys ala ala ser glu val ala arg thr val gly ser val ala lys ser met gly asp tyr  
841/281 871/291  
TTG GAT TCA GAG GCA GAG ACC AAT GAG ATC AAG ACC GTC GTC TCG GAG GAG TCA GAG  
leu asp ser his pro ala thr thr thr val met thr ala val leu gln ala ala val gly  
901/301 931/311  
TTG GGT TCA GTC GCA TCA CTG AAG GCG GAT TTT GAG GCG AAT GCG GAG GAG TCA GAG  
pro gly ser val ala ser leu lys ala his phe ala ala pro lys val ala ser asp

SEQ ID NO: 1A'

[illegible]

TCG CCT TTG ACG CCT ATT CGG TCG CGC AGC TTT TTG GCG ACG TCA CTG GTG CCC GCG TTG  
ser pro leu thr pro ile arg ser arg ser phe leu ala thr ser leu val pro ala leu  
63/21

CGG GCG TCC AGC CGC AGC GAC ACC ACA TAC GGC CGS TCC GGT TGC GGG GGC CGT TGG GTG  
arg ala ser ser arg ser asp thr thr tyr gly arg ser gly cys gly gly arg trp val  
123/41

GGG TTG GGT GCC TCC GTC ACC CCA GGC AGT TCG CTG GCT ATT TGT CGC AGT AGC GCG ACG  
gly leu gly gly ala ser val thr pro gly ser ser leu ala ile cys arg ser ser ala thr  
183/61

GGA TTG TCG ATG TCT TGG TAG CTA GCA TCC GGT CGG GGG GCC GCT ACC AGC GCC AGC GCC  
ala leu ser met ser trp AMB leu ala ser gly arg gly ala ala thr ser ala ser ala  
243/81

GSG GCT CCC CGG TCC GGG TAS TGC GCG TCG AGT TGG TCG TGG ACC AGC AAT GAC TGC GAC  
gly ala pro arg ser gly AMB cys ala ser ser trp ser trp thr ser asn asp cys asp  
303/101

CGG GCG ACT TCG AAA CCG CCA CCG GTT ACA TTC CGG GAG TGC CTC ATC GCC AGG TAA ACC  
pro ala thr ser lys pro pro pro val asg phe pro asp cys val ile ala arg OCH thr  
363/121

GCC GGC ACT AAC GCC AGC AAC CAA CCG GTG AAG ACC AAT AAA CGG CAC CTG CGC ASS TTG  
ala gly thr asn ala ser asn gln pro val lys thr asn gln arg his leu arg arg leu  
423/141

CGG CTC AAC CGC ATC ATG AAC TGC TGS ATT TCG GAG TCG CGG TAG TCT CGC GCA GTG CGT  
arg leu asn arg ile met asn cys trp ile ser asp ser pro tyr ser arg ala val arg  
483/161

GCC CGC GAG CCT ACC GAA GAT GGC GTG CAT GCS TTC GSC GTS GAC CGC ACA GCA CCT GGA  
ala arg glu pro thr glu asp arg val his ala phe gly val asp arg thr ala pro gly  
543/181

GTT GGC GGC GCC GAG GGC CGA GAT GGC AGG ATT AGC GAU CMT CGS GCG CGG GAA CTC CCA  
val gly gly ala glu gly arg asp gly arg met thr asp arg arg gly arg glu leu pro  
603/201

GSC CGC CGS ACC GTG GCA AAU CCS TCS CAA ACT PRT LIT AAA CSB TAA GGA CTC ATI SAT  
gly arg arg thr val ala asp pro ser gln thr ala thr lys pro OCH gly val ile pro  
663/221

GAA GAC ASS CAC CGC GAG GAI GCS SSS CAG NST NIT LI AAT AIT GAT CGS COT HSI STI  
glu asp arg his arg asp asp ala ala gln ala val gly ser thr asp arg pro arg val  
723/241

TCC GGS GCG CSC CST TCG GTC GTC TCG CAA ACT AIT AAT AAT AAT AAT AAT AAT AAT AAT  
ala gly gly ala arg tyr thr ala gly arg thr thr ser asp asp arg val gly pro val  
783/261

LHU GSB CAI CVA ANT BUI GAI SA LA TI AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg gly thr arg asp ile thr asp gly arg thr thr ser asp asp arg val gly pro val  
843/281

SRA TTS ACA SST AGA WLI YAA YIA IUT SAT SA ITI AIT AIT AIT AIT AIT AIT AIT AIT  
gly phe thr pr arg asp ala pr gly IPT IST IIS IIS IIS IIS IIS IIS IIS IIS  
903/301

HIL DLT TST CHL ADP RLA GAI MHI HIT DLI DLI DLI DLI DLI DLI DLI DLI DLI DLI  
his ile thr chl adp rla gai mhi hit dli dli dli dli dli dli dli dli dli dli  
963/321

DIV DLT HIL DLT DIV DLT DIV DLT DIV DLT DIV DLT DIV DLT DIV DLT DIV DLT DIV DLT  
div dlt hil dlt div dlt div dlt div dlt div dlt div dlt div dlt div dlt div dlt  
1023/341

5/185

Séquence codante DP428 identique à la séquence Rv0203 prédite par Cole et al.  
(Nature 393:537-544)

```

1/1                               31/11
ATG AAG ACA GGC ACC GCG ACG ACG CGG CGC AGG CTG TTG GCA GTA CTG ATC GCC CTC GCG
Met lys thr gly thr ala thr thr arg arg arg leu leu ala val leu ile ala leu ala
61/21                               91/31
TTG CCG GGG GCC GCC GTT GCG CTG CTG GCC GAA CCA TCA GCG ACC GGC GCG TCG GAC CCG
leu pro gly ala ala val ala leu leu ala glu pro ser ala thr gly ala ser asp pro
121/41                               151/51
TGC GCG GCC AGC GAA CTG GCG AGG ACG GTC GGT TCG GTC GCC AAG TCG ATG GCC GAC TAC
cys ala ala ser glu val ala arg thr val gly ser val ala lys ser met gly asp tyr
181/61                               211/71
CTG GAT TCA CAC CCA GAG ACC AAC CAG GTG ATG ACC GCG GTC TTG CAG CAG CAG GTA GCG
leu asp ser his pro glu thr asn gln val met thr ala val leu gln gln gln val gly
241/81                               271/91
CCG GGG TCG GTC GCA TCG CTG AAG GCC CAT TTC GAG GCG AAT CCC AAG GTC GCA TCG GAT
pro gly ser val ala ser leu lys ala his phe glu ala asn pro lys val ala ser asp
301/101                               331/111
CTG CAG GCG CTT TCG CAA CCG CTG ACC GAT CTT TCG ACT GCG TCG TCG CTG CCG ATC AGC
leu his ala leu ser gln pro leu thr asp leu ser thr arg cys ser leu pro ile ser
361/121                               391/131
GGT CTG CAG GCG ATC GGT TTG ATG CAG GCG CTG CAG GCG GCG GCG GCG TAG
gly leu gln ala ile gly leu met gln ala val gln gly ala arg arg AMB

```

SEQ ID N° 1D

FIGURE 1D

ORF contenant la séquence DP428 et faisant partie de seq1A'

```

1/1                               31/11
TGA GGG ATC GTC GGG GGT GGG AAC TCG CAG GTC GTC GCA GCG TCG CAA ACC CGT GCG AAA
GPA arg ile val gly gly gly asn ser gln ala ala gly pro ser gln thr arg arg lys
61/21                               91/31
CCG GTC GCA AAC GGT AAG CAG TCA TCG ATG AAG ACA GTC AAT GCG ATG ACC GCG GCG AGC
pro val ala asn arg lys gln ser ser met lys thr gly thr ala thr thr arg arg arg
121/41                               151/51
CTG TTG GCA GTA CTG ATC GCG GTC GCG TTT TCG GAG GTC GCG GTT GCG CTG CTG GCG GAA
leu leu ala val leu ile ala leu ala leu pro gly ala ala val ala leu leu ala gln
181/61                               211/71
GCA TCA GCG AAT GAT GTC TCG GAT GTC TCG GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
pro ser ala thr gly ala ser arg pro gly ala ala ser ala thr thr thr thr thr thr
241/81                               271/91
TTG CTG GGT AAG TCG AAT GTC GAT TAT TTT GTC TCA GTC GTC GTC GTC GTC GTC GTC GTC
leu val ala ly ser pro gly asp thr leu arg ser ser pro gln thr asn gln val pro
301/101                               331/111
ATG GTC GTC TCG TCG GAG GAG GAA GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
thr ala val leu ala ala ala val gly ser gly ser ser ala thr thr thr thr thr thr
361/121                               391/131
GAT GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
asp thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr
421/141                               451/151
GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
val thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

```

SEQ ID N° 1D

6/185

SEQ. IN. N. 1.

7/185

31/11  
 TCG CCG GCT CGC GGA CGT AGA TAA TAG CTC ACC GTT GGA CGA CCT CGA CAG GGT CCT TTG  
 ser pro ala arg gly arg arg OCH AMB leu thr val gly arg pro arg gln gly pro leu  
 61/21  
 91/31  
 TGA CTG CCG GGC TTG ACG CGG ACG ACC ACA GAG TCG GGT CAT CGC CTA AGG CTA CCG TTC  
 OPA leu pro gly leu thr arg thr thr thr glu ser gly his arg leu arg leu pro phe  
 121/41  
 151/51  
 TGA CCT GGG GTG CGT GGG CGC CGA CGA GTG AGG CAG TCA TGT CTC AGG GCC CAC CGC CAC  
 OPA pro gly val arg gly arg arg arg val arg gln ser cys leu arg ala his arg his  
 181/61  
 211/71  
 CTC GGT CGC CGG CAG TGT CAG CAT GTG CAG ATG ACT CCA CGC AGC TTG TTC GTG TTG GTG  
 leu gly arg arg gln cys gln his val gln met thr pro arg ser leu phe val leu val  
 241/81  
 271/91  
 TCG TGG TTG CGA CGA CTT GGC GCT GGT GAG CGC ACC CGC CGG CGT CGT GCC GCG CAT GCG  
 ser trp leu arg arg leu gly ala gly glu arg thr arg arg arg arg ala ala his ala  
 301/101  
 GAT C  
 asp

SEQ ID N° 3A

FIGURE 3A

32/11  
 CGC CGG CTC GCG GAG GTA GAT AAT AGC TCA CCG TTG GAT GAG CTC GAG AGG CTC CTT TGT  
 arg arg leu ala asp val asp asp ser ser pro leu asp asp leu asp arg val leu cys  
 62/21  
 92/31  
 GAG TGG CCG GGT TGA CGC GGA CGA CCA CAG AGT CGG GTT ATT GGC TAA GGT TAT TGT TTT  
 asp cys arg ala GHA arg gly arg pro gln ser arg val ala ala OCH gly tyr arg ser  
 102/41  
 132/51  
 GAG CTC GGT TGC CTC GAT GAT GAT GAT TGA GAT AGT GAT GAT TGA GAT TAT AGT GAT GAT  
 asp leu gly cys val gly ala asp ala GHA gly ser leu val ser gly pro thr ala thr  
 162/61  
 192/71  
 TGT GTT TAT GAT GAT TTT GGT GAT GAT GAT GAT TTA GAT TAT GAT TAT TAT TAT TAT TAT  
 thr val thr gly ser val ser met thr arg ala leu asp ala ala cys thr thr thr thr  
 222/81  
 252/91  
 GGT GGT TAT GAT GAT TTT GGT GAT GAT GAT GAT TTA GAT TAT GAT TAT TAT TAT TAT TAT  
 asp gly thr asp arg leu ala leu val ser ala asp ala gly val val pro thr thr thr  
 282/101  
 GAT C  
 asp

33/11

30

FIGURE 3C

[illegible]



9/185

CAA TTT TCC TTC GCG CCG TGC AAT ACC ATC TGC AAG ACC AGC GAC GGC CCG TGG TTG CCG  
gln phe ser phe ala pro cys asn thr ile cys lys thr ser asp gly pro trp leu arg  
62/21 92/31

TGC CGC AGC TTG CGG AAA CCG GGT ATG CAC CCT GCC GTA CCG TTG TTG CCA CTT GAT GTC  
ser arg ser leu arg lys pro gly met asp pro ala val pro leu leu pro leu asp val  
122/41 152/51

GTC GCT CTC CAC CCG TCG GGG GGC GAA AGC CAT TCC GAC ACT GGG ATC CTC AAA ACG TCG  
val ala leu his pro ser gly gly glu ser his ser asp thr gly ile leu lys thr ser  
182/61 212/71

GCT GAG TGT CTG CAG GGC TCC GGG GAG CAG CCG ATC ATC ACC ATG TAC GAA CTG AAT AAG  
ala glu cys leu gln gly ser gly glu gln pro ile ile thr met tyr glu leu asn lys  
242/81 272/91

TCC CCC GCG CGC GAC TTC CAG ACA TTT GTT GTG GTT TGG GTT GAG GCC GAG CCG AGG CTC  
ser pro ala arg asp phe gln thr phe val val val ser val glu ala glu ala arg leu  
302/101 332/111

ATT TCG CAG CAA CCG GTC TCC GGG TCG CAG CAT GGT TCG GCG GAT CGC GGC GCA GTC GTC  
ile ser gln gln pro val ser gly ser gln his arg cys gly asp arg gly ala val val  
362/121

GGA CGA GTC GTC GTC AAC GAC CAC GAT C  
gly arg val val val asn asp his asp

SEQ ID N° 4B

FIGURE 4B

[illegible]

1000

## STATEMENT OF WORK

10/185

partie de la séquence nucléotidique de seq4A

```

1/1                               31/11
CCG CGC GCG ACT TCC AGA CAT TTG TTG TGG TTT CGG TTG AGG CCG AGG CGA GGC TCA TTT
pro arg ala thr ser arg his leu leu trp phe arg leu arg pro arg arg gly ser phe
61/21                               91/31
CGC AGC AAG CCG TCT CCG GGT CGC AGC ATC GTT GCG GCG ATC GCG GCG CAG TCG TCG GAC
arg ser lys arg ser pro gly arg ser ile val ala ala ile ala ala gln ser ser asp
121/41
GAG TCG TCG TCA ACG ACC ACG ATC
glu ser ser ser thr thr thr ile

```

SEQ ID N° 4A'

FIGURE 4A'

```

1/1                               31/11
CGC GCG CGA CTT CCA GAC ATT TGT TGT GGT TTC GGT TGA GGC CGA GGC GAG GCT CAT TTC
arg ala arg leu pro asp ile cys cys gly phe gly OPA gly arg gly glu ala his phe
61/21                               91/31
GCA GCA AGC GGT CTC CGG GTC GCA GCA TCG TTG GCG CGA TCG CCG CGC AGT CGT CGG ACG
ala ala ser gly leu arg val ala ala ser leu arg arg ser arg arg ser arg thr
121/41
AGT CGT CGT CAA CGA CCA CGA TC
ser arg arg gln arg pro arg

```

SEQ ID N° 4B'

FIGURE 4B'

```

1/1                               31/11
CGC GCG CGA CTT CCA GAC ATT TGT TGT GGT TTC GGT TGA GGC CGA GGC GAG GCT CAT TTC
arg ala arg leu pro asp ile cys cys gly phe gly OPA gly arg gly glu ala his phe
61/21                               91/31
GCA GCA AGC GGT CTC CGG GTC GCA GCA TCG TTG GCG CGA TCG CCG CGC AGT CGT CGG ACG
ala ala ser gly leu arg val ala ala ser leu arg arg ser arg arg ser arg thr
121/41
AGT CGT CGT CAA CGA CCA CGA TC
ser arg arg gln arg pro arg

```

FIGURE 4C'

FIGURE 4D' (N° 4E' et 4F')

11/185

ORF d'après par Cole et al. (Nature 393:537-544) et contenant la séquence Seq 4A'

```

1/1                               31/11
tga ata agt ccg ccg cgc gcg act tcc aga cat ttg ttg tgg ttt cgg ttg agg ccg agg
OPA ile ser pro pro arg ala thr ser arg his leu leu trp phe arg leu arg pro arg
61/21                               91/31
cga ggc tca ttt cgc agc aag ccg tct ccg ggt cgc agc atc gtt gcg gcg atc gcg gcg
arg gly ser phe arg ser lys arg ser pro gly arg ser ile val ala ala ile ala ala
121/41                               151/51
cag tcg tcg gac gag tcg tcg tca acg acc acg atc tcg aac tcg acg ccc tcc tgt tcg
gln ser ser asp glu ser ser ser thr thr thr ile ser asn ser thr pro ser cys ser
181/61                               211/71
agg atg cta cgc aga cag cgc tcg atg gtg gcg ccg ttg ttg tac atc ggg atg cac acc
arg met leu arg arg gln arg ser met val ala pro leu leu tyr ile gly met his thr
241/81                               271/91
gag ata agc ggt ttc gcc ggg ttc acc gat acc acg ctt gat gca tca cca gcc acc acc
gln ile ser gly phe ala gly phe thr asp thr thr leu asp ala ser pro gly thr thr
301/101
tgg cga ctc aga gac tag
trp arg leu arg asp AMB

```

SEQ ID N° 4F

FIGURE 4F

séquence en amont de seq4A' et en fusion avec seq4A'

```

1/1                               31/11
GCA ACC TAC CAG CAG AGC CAG GGG CTC ACA GGA CCT AAA GGA GTA GCG CCC ATC CCT GAT
ala thr tyr gln gln ser gln gly leu thr gly pro lys gly val ala pro met ala asp

```

C

SEQ ID N° 4J

FIGURE 4J

seq4J' dans une autre phase de lecture

```

1/1                               31/11
ACG CAA CTT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT
thr gln pro thr ser ala ala thr gly ser ala arg leu tyr gln asp thr pro thr thr
61/21
ATC
ala

```

SEQ ID N° 4K

FIGURE 4K

SEQ ID N° 4L

FIGURE 4L

12/185

séquence Rv2050 prédite par Cole et al. (Nature 393:537-544) et contenant seq4J  
 1/1 31/11  
 ATG GCT GAT CGT GTC CTG AGG GGC AGT CGC CTC GGA GCC GTG AGC TAT GAG ACC GAC CGC  
 Met ala asp arg val leu arg gly ser arg leu gly ala val ser tyr glu thr asp arg  
 61/21 91/31  
 AAC CAC GAC CTG GCG CCG CGC CAG ATC GCG CGG TAC CGC ACC GAC AAC GGC GAG GAG TTC  
 asn his asp leu ala pro arg gln ile ala arg tyr arg thr asp asn gly glu glu phe  
 121/41 151/51  
 GAA GTC CCG TTC GCC GAT GAC GCC GAG ATC CCC GGC ACC TGG TTG TGC CGC AAC GGC ATG  
 glu val pro phe ala asp asp ala glu ile pro gly thr trp leu cys arg asn gly met  
 181/61 211/71  
 GAA GGC ACC CTG ATC GAG GGC GAC CTG CCC GAG CGG AAG AAG GTT AAG CCG CCC CGG ACC  
 glu gly thr leu ile glu gly asp leu pro glu pro lys lys val lys pro pro arg thr  
 241/81 271/91  
 CAC TGG GAC ATG CTG CTG GAG GCG CGT TCC ATC GAA GAA CTC GAA GAG TTA CTT AAG GAG  
 his trp asp met leu leu glu arg arg ser ile glu glu leu glu glu leu leu lys glu  
 301/101 331/111  
 CGC CTC GAG CTC ATT CGG TCA CGT CGG CGC GGC TGA  
 arg leu glu leu ile arg ser arg arg arg gly cys

SEQ ID N° 4M

FIGURE 4M

ORF d'après par Cole et al. Nature 393:537-544 et contenant la séquence Rv2050  
 1/1 31/11  
 TAG TCC GGC CGG GTG TCC GAT GCG GGT ATC ATT GAT GAT GCG GGC GCG CGC GTC GCG TCG  
 AMB ser ala arg val ser asp pro gly ile ile asp ile arg ala ala arg val ala cys  
 61/21 91/31  
 CGG GAA CTA CGC ACA CGG CCG GAG CGT TTS CCA A GAA ACC AAT CCG GAG TAC CCA ACC  
 arg glu leu arg arg arg pro gln arg leu pro thr gly ala ser arg gln tyr ala thr  
 121/41 151/51  
 TAC CAG CAG AAT CCA GCG CTC ACA GGA CTT AAA GGA GGA GTC CTT ATG GGT GAT GGT GTC  
 tyr gln gln ser pro gly leu thr gly pro lys gly val ala pro thr ala asp arg val  
 181/61 211/71  
 CTC AAG GAT AAT TAT TTT GAA GAT CTG AAT GAT GAA A GAT GAA GAT GAA GAT GAT GAT  
 leu arg gly ser arg leu gln ile val ser thr gly thr thr thr thr thr thr thr thr  
 241/81 271/91  
 GAT CCA GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 pro arg ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala  
 301/101 331/111  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 asp asp ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala  
 361/121 391/131  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 421/141 451/151  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 481/161 511/171  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 541/181 571/191  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 601/201 631/211  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 661/221 691/231  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 721/241 751/251  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 781/261 811/271  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 841/281 871/291  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 901/301 931/311  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 961/321 991/331  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1021/341 1051/351  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1081/361 1111/371  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1141/381 1171/391  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1201/401 1231/411  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1261/421 1291/431  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1321/441 1351/451  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1381/461 1411/471  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1441/481 1471/491  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1501/501 1531/511  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1561/521 1591/531  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1621/541 1651/551  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1681/561 1711/571  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1741/581 1771/591  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1801/601 1831/611  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1861/621 1891/631  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1921/641 1951/651  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 1981/661 2011/671  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2041/681 2071/691  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2101/701 2131/711  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2161/721 2191/731  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2221/741 2251/751  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2281/761 2311/771  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2341/781 2371/791  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2401/801 2431/811  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2461/821 2491/831  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2521/841 2551/851  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2581/861 2611/871  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2641/881 2671/891  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2701/901 2731/911  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2761/921 2791/931  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2821/941 2851/951  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2881/961 2911/971  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 2941/981 2971/991  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 3001/1000  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr

SEQ ID N° 4M

FIGURE 4M

13/185

31/11  
 GAT CGC GGT CAA CGA GGC CGA ATA CGG CGA GAT GTG GGC CCA AGA CGC CGC CGC GAT GTT  
 asp arg gly gln arg gly arg ile arg arg asp val gly pro arg arg arg arg asp val  
 61/21  
 TGG CTA CGC CGC GGC GAC GGC GAC GGC GAC GGC GAC GTT GCT GCC GTT CGA GGA GGC GCC  
 trp leu arg arg gly asp gly asp gly asp gly asp val ala ala val arg gly gly ala  
 121/41  
 GGA GAT GAC CAG CGC GGG TGG GCT CCT CGA GCA GGC CGC CGC GGT CGA GGA GGC CTC CGA  
 gly asp asp gln arg gly trp ala pro arg ala gly arg arg gly arg gly gly leu arg  
 181/61  
 CAC CGC CGC GGC GAA CCA GTT GAT GAA CAA TGT GGC GCA GGC GCT GCA ACA GCT GGC CCA  
 his arg arg gly glu pro val asp glu gln cys ala pro gly ala ala thr ala gly pro  
 241/81  
 GCC CAC GCA GGC CAC CAC GGC TTC TTC CAA GGT GGC TGG CCT CTG GAA GAC GGT CTC GCC  
 ala his ala gly his his ala phe phe gln ala gly trp pro val glu asp gly leu ala  
 301/101  
 GCA TCG GTC GCG GAT C  
 ala ser val ala asp

SEQ ID N° 5A

FIGURE 5A

32/11  
 ATC GCG GTC AAC GAG GCG GAA TAC GGC GAG ATC TGG GCG GAA GAC GCG GCG GCG ATG TTT  
 ile ala val asn glu ala glu tyr gly gln met trp ala glu asp ala ala ala met phe  
 62/21  
 GGC TAC GCG GCG GCG ACG GCG ACG GCG ACG GCG ACG TTT CTT CTT CTT CTT GAG GAG GCG GCG  
 gly tyr ala ala ala thr ala thr ala thr ala thr ala thr ala thr ala thr ala thr ala thr  
 122/41  
 GAG ATG ACG ACG GCG GGT GCG CTC CTC GAG CAG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG  
 glu met thr ser ala gly gly leu leu glu glu ala ala ala val glu glu ala ser asp  
 182/61  
 ACG GCG GCG GCG AAG CAG TTC ATC AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG AAG  
 thr ala ala ala asn gln leu met ala asp thr pro glu ala leu thr glu leu ala glu  
 242/81  
 TTT AAT AAT AAT AAT AAT TTT TTT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
 pro thr ala gly thr thr pro thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
 302/101  
 GAT GCG TCG GAT ATC  
 his arg arg pro ile

SEQ ID N° 5B

14/185

33/11  
 TCG CGG TCA ACG AGG CCG AAT ACG GCG AGA TGT GGG CCC AAG ACG CCG CCG CGA TGT TTG  
 ser arg ser thr arg pro asn thr ala arg cys gly pro lys thr pro pro arg cys leu  
 63/21  
 GCT ACG CCG CCG CGA CGG CGA CGG CGA CGG CGA CGT TGC TGC CGT TCG AGG AGG CGC CGG  
 ala thr pro arg arg arg arg arg arg arg arg arg cys cys arg ser arg arg arg arg  
 123/41  
 AGA TGA CCA GCG CCG GTG GGC TCC TCG AGC AGG CCG CCG CCG TCG AGG AGG CCT CCG ACA  
 arg opa pro ala arg val gly ser ser ser arg pro pro arg ser arg arg pro pro thr  
 183/61  
 CCG CCG CCG CGA ACC AGT TGA TGA ACA ATG TGC CCC AGG CGC TGC AAC AGC TGG CCC AGC  
 pro pro arg arg thr ser opa opa thr met cys pro arg arg cys asn ser trp pro ser  
 243/81  
 CCA CGC AGG GCA CCA CGC CTT CTT CCA AGC TGG GTG GCC TGT GGA AGA CGG TCT CGC CGC  
 pro arg arg ala pro arg leu leu pro ser trp val ala cys gly arg arg ser arg arg  
 303/101  
 ATC GGT CGC CGA TC  
 ile gly arg arg

SEQ ID N° 5C

FIGURE 5C

partie de la séquence nucléotidique Seq 5A

1/1 31/11  
 GGC GGC GGC GAC GGC GAC GGC GAC GGC GAC GTT GGT GGC GTT CCA GGA GGC GGC GGA GAT  
 arg arg gly asp gly asp gly asp gly asp val ala ala val arg gly gly ala gly asp  
 61/21  
 GAC CAG CCG CCG TGG GCT CCT CGA GCA GGC CGC CGC GGT CGA GGA GGT GTC CGA GAC CTC  
 asp gln arg gly trp ala pro arg ala gly arg arg gly arg gly gly leu arg his arg  
 121/41  
 GGC GGA GAA GGA GTT GAT GAA GAA TGT GGT GGA GGC GGT GCA AAT GAT GAT GCA GAT GAT  
 arg gly ala pro val arg ala gln trp ala pro gly ala ala thr ala gly pro ala his  
 181/61  
 GGA GGA GAT GAT GGT TTT TTT GAA GGT GAT TAT GAT GAT GAA GTT GAT GAT GAT GAA GAT  
 ala gly asp ala ala pro pro ala ala gly trp pro val ala arg gly leu ala ala arg  
 241/81  
 GTC GAT GAT  
 val ala arg

15/185

```

1/1                               31/11
TAC GCC GCG GCG ACG GCG ACG GCG ACG GCG ACG TTG CTG CCG TTC GAG GAG GCG CCG GAG
tyr ala ala ala thr ala thr ala thr ala thr leu leu pro phe glu glu ala pro glu
61/21                               91/31
ATG ACC AGC GCG GGT GGG CTC CTC GAG CAG GCC GCC GCG GTC GAG GAG GCC TCC GAC ACC
met thr ser ala gly gly leu leu glu gln ala ala ala val glu glu ala ser asp thr
121/41                               151/51
GCC GCG GCG AAC CAG TTG ATG AAC AAT GTG CCC CAG GCG CTG CAA CAG CTG GCC CAG CCC
ala ala ala asn gln leu met asn asn val pro gln ala leu gln gln leu ala gln pro
181/61                               211/71
ACG CAG GGC ACC ACG CCT TCT TCC AAG CTG GGT GGC CTG TGG AAG ACG GTC TCG CCG CAT
thr gln gly thr thr pro ser ser lys leu gly gly leu trp lys thr val ser pro his
241/81
CGG TCG CCG ATC
arg ser pro ile

```

SEQ ID N° 5B'

FIGURE 5B'

```

1/1                               31/11
ACG CCG CGG CGA CGG CGA CGG CGA CGG CGA CCT TCG TCG CCT TCG AGG AGG CGG CGG AGA
thr pro arg arg arg arg arg arg arg arg arg cys cys arg ser arg arg arg arg
61/21                               91/31
TGA CCA CGG CGG GTG GGC TCG TCG AGC AGG CCG CGG CGG TCG AGG AGG CCT CGG ACA CGG
OPA pro ala arg val gly ser ser ser arg pro pro arg ser arg arg pro pro thr pro
121/41                               151/51
CCG CGG CGA ACC AGT TGA TGA ACA ATG TCG CCG AGG CGC TCG AAC AGC TGG CCG AGG CCA
pro arg arg thr ser OPA OPA thr met cys pro arg arg cys asn ser trp pro ser pro
181/61                               211/71
CGC AGG GCA CCA CGG CTT CTT CCA AAT TCG CTG GCG TGT CCA ACA CCG TGT CCG CCG AGC
arg arg ala pro arg leu leu thr ser trp val ala cys gly arg arg ser arg arg ile
241/81
GCT TGT CCA CCA
gly arg arg

```

SEQ ID N° 5C'

FIGURE 5C'

16/185

ORF prédite par Cole et al. (Nature 393:537-544) et contenant seq5A'

1/1 31/11

tga act gat gat tct gat agc gac caa cct ctt ggg gca aaa cac ccc ggc gat cgc ggt  
 OPA thr asp asp ser asp ser asp gln pro leu gly ala lys his pro gly asp arg gly  
 61/21 91/31

caa cga ggc cga ata cgg cga gat gtg ggc cca aga cgc cgc cgc gat gtt tgg cta cgc  
 gln arg gly arg ile arg arg asp val gly pro arg arg arg arg asp val trp leu arg  
 121/41 151/51

cgc ggc gac ggc gac ggc gac ggc gac gtt gct gcc gtt cga gga ggc gcc gga gat gac  
 arg gly asp gly asp gly asp gly asp val ala ala val arg gly gly ala gly asp asp  
 181/61 211/71

cag cgc ggg tgg gct cct cga gca ggc cgc cgc ggt cga gga ggc ctc cga cac cgc cgc  
 gln arg gly trp ala pro arg ala gly arg arg gly arg gly leu arg his arg arg  
 241/81 271/91

ggc gaa cca gtt gat gaa caa tgt gcc cca ggc gat gca aca gct ggc cca gcc cac gca  
 gly glu pro val asp glu gln cys ala pro gly ala ala thr ala gly pro ala his ala  
 301/101 331/111

ggg cac cac gcc ttc ttc caa gct ggg tgg cct atg gaa gac ggt ctc gcc gca tgc gtc  
 gly his his ala phe phe gln ala gly trp pro val glu asp gly leu ala ala ser val  
 361/121 391/131

gcc gat cag caa cat ggt gtc gat ggc caa caa cca cat gtc gat gac caa ctc ggg tgt  
 ala asp gln gln his gly val asp gly gln gln pro his val asp asp gln leu gly cys  
 421/141 451/151

gtc gat gac caa cac ctt gaq ctc gat gtt gaa ggg ctt tgc tgc ggc ggc ggc cgc cca  
 val asp asp gln his leu glu leu asp val glu gly leu cys ser gly gly gly arg pro  
 481/161 511/171

ggc cgt gca aac cgc ggc gca aaa cgg ggt cgc ggc gat gag ctc gct ggg cag ctc gct  
 gly arg ala asn arg gly ala lys arg gly pro gly asp glu leu ala gly gln leu ala  
 541/181 571/191

ggg ttc ttc ggg tct ggg cgg tgg ggt gcc cca caa ctt gaa tgg ggc ctc ggt cgg  
 gly phe phe gly ser gly arg trp gly gly arg ala leu gly ser gly gly leu gly arg  
 601/201 631/211

ttc gtt gtc ggt gcc gca ggc ctg ggc cgc gac caa cca gac ggt cag ccc ggc ggc ggc  
 phe val val gly ala ala gly leu gly arg gly ala pro gly ser his pro gly gly ala  
 661/221 691/231

ggc gct gcc gct gac cag cct gat cag cca ctt gaa agc gaa gaa cgc gca gat gct ggc  
 gly ala ala ala asp gln pro asp gln arg arg gly lys arg ala arg ala asp ala gly  
 721/241 751/251

cag gct gcc ggt ggc gca gac ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc  
 arg ala ala gly gly ala asp gly arg gln gly ala trp trp ala gln trp cys ala ala  
 781/261 811/271

gac  
 cys ser ala ala thr leu cys asp ala ala pro ser gly ala thr leu gly ala gly ala  
 841/281

cag act gac gac gac gac gac gac gac gac gac gac gac gac gac gac gac gac gac gac gac  
 gln thr val val ala OPA

SEQ ID NO: 1



séquence Rv1196 prédite par Cole et al. (Nature 393:537-544) et pouvant coder pour une ORF en fusion avec Seq5A'

17/1  
atg gtg gat ttc ggg gcg tta cca ccg gag atc aac tcc gcg agg atg tac gcc gcc ccg  
Met val asp phe gly ala leu pro pro glu ile asn ser ala arg met tyr ala gly pro  
391/21  
ggg tgg gcc tgg ctg gtg gcc gcg gct cag atg tgg gac agc gtg gcg agt gac ctg ttt  
gly ser ala ser leu val ala ala ala gln met trp asp ser val ala ser asp leu phe  
121/41  
tcg gcc gcg tgg gcg ttt cag tgg gtg gtc tgg ggt ctg acg gtg ggg tgg tgg ata ggt  
ser ala ala ser ala phe gln ser val val trp gly leu thr val gly ser trp ile gly  
181/61  
tcg tgg gcg ggt ctg atg gtg gcg gcg gcc tgg ccg tat gtg gcg tgg atg agc gtc acc  
ser ser ala gly leu met val ala  
ala ala ser pro tyr val ala trp met ser val thr  
241/81  
ggc ggg cag gcc gag ctg acc gcc gcc cag gtc ccg gtt gct gcg gcg gcc tac gag acc  
ala gly gln ala glu leu thr ala ala gln val arg val ala ala ala ala tyr glu thr  
301/101  
ggc tat ggg ctg acg gtg ccc ccg ccg gtg atc gcc gag aac cgt gct gaa ctg atg att  
ala tyr gly leu thr val pro pro pro val ile ala glu asn arg ala glu leu met ile  
361/121  
ctg ata gcg acc aac ctc ttg ggg caa aac acc ccg gcg atc gcg gtc aac gag gcc gaa  
leu ile ala thr asn leu leu gly gln asn thr pro ala ile ala val asn glu ala glu  
421/141  
tac gcc gag atg tgg gcc caa gac gcc gcc gcg atg ttt gcc tac gcc gcg gcg acc gcg  
tyr gly glu met trp ala gln asp ala ala ala met phe gly tyr ala ala ala thr ala  
481/161  
acg gcg acg gcg acg ttg ctg ccg ttc gag gag gcg ccg gcg gag atg acc agc gcg ggt ggg  
thr ala thr ala thr leu leu pro phe glu glu ala pro glu met thr ser ala gly gly  
541/181  
ctc ctc gag cag gcc gcc gcg gtc gag gag gcc tcc gac acc gcc gcg gcg aac cag ttg  
leu leu glu gln ala ala ala val glu glu ala ser asp thr ala ala ala asn gln leu  
601/201  
atg aac aat gtg ccc cag gcc ctg caa cag ctg gcc cag acc acc gcg ggg acc acc cct  
met asn asn val pro gln ala leu gln gln leu ala gln pro thr gln gly thr thr pro  
661/221  
tat tcc aag atg ggt gcc ctg tgg aag acg gtc tgg tgg cat cct tgg gcc acc acc acc  
ser ser lys leu gly gly leu trp lys thr val ser pro ala arg ser pro ala ser asn  
721/241  
atg gtg tgg atg acc aac aac caa atg tgg atg atg aac tgg ggt ggt tcc atg acc aac  
met val ser met ala asn asn his met ser met thr asn ser gly val ser met thr asn  
781/261  
aaa tgg aga tgg atg tgg aac gcc ttt gcc gcg gac gac acc acc acc acc acc acc  
thr leu ser ser met leu lys gly phe ala pro ala ala ala ala ala ala val ala thr  
841/281  
gag gcc caa caa ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt  
ala ala acc gln gly val ala ala met acc ser leu gln gln gln gln gln gln gln gln  
901/301  
gag ggt ttt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt  
leu gly gly val ala ala ala acc leu gly ala ala ala ala ala ala ala ala ala ala  
961/321  
gag atg ggt tgg tgg ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt  
pro gln ala trp ala ala ala ala ala ala ala val thr thr ala ala ala ala ala ala  
1021/341

Seq 5P: ORF d'après Cole et al. (Nature 393.537-544) et contenant la séquence Rv1196

DECLASSIFICATION AUTHORITY: 25X

19/185

31/11  
 GGA TCC TGA TGC AAG TGG TCC GGG ATT TGT CGG CAG CCA CCG CGG TCC CGT CGA CCA ACG  
 gly ser OPA cys lys trp ser gly ile cys arg gln pro arg arg ser arg arg pro thr  
 61/21  
 TTG GTG CAT CCG GGC TGC GAG CAT GCA CGC ACC GAC CAG CGC GGC GAG CGC GGC TAG CTG  
 leu val his pro gly cys glu his ala arg thr asp gln arg gly glu arg gly AMB leu  
 121/41  
 CTT GCC CAC TGT TCC TCC CTG CCG GCA CCA TGT GCG ACA AGC TTA AGC GCA GCA GTA CCG  
 leu ala his cys ser ser leu pro ala pro cys ala thr ser leu ser ala ala val pro  
 181/61  
 CCG GTG CCT GGG CAT CCA GCA AAA CGG GGA GCF CAA GAA CGA TTC ATG AAC GAG GGG TCG  
 ala val pro gly his pro ala lys arg gly ala gln glu arg phe met asn glu gly sec  
 241/81  
 TCA CCA ACG TCG AAA CCG ACG GTT GCC AGC CGG CCC ACG ATA TTG CGT GGT CGA GGG TCC  
 ser pro thr ser lys pro thr val ala ser arg pro thr ile leu arg ala arg gly ser  
 301/101  
 SCT GTA CCC TCA CCS AAC GTG AGT CCC ACA CCG CGG AGG CGG GCG ACT CTG GCG TCG TTA  
 ala val pro ser pro asn val ser pro thr pro arg arg arg ala thr leu ala ser leu  
 361/121  
 GCA GCC GAG CTC AAG GTS TCC CGC ACC ACT GTC TCG AAT CCT TTT AAC CGA CCG GAT CGA  
 ala ala glu leu lys val ser arg thr thr val ser asn ala phe asn arg pro asp pro  
 421/141  
 GAA GGA GAA GAT C  
 glu gly glu asp

SEQ ID N° 6A

FIGURE 6A

32/11  
 GAT CCT GAT GGA ACT GGT CCG GGA TTT GTC GGC AGC GAG GGC GGT CCG GTC GAG CAA CGT  
 asp pro asp ala ser gly pro gly phe val gly ser his gly gly pro val asp gln arg  
 62/21  
 TGG TGG ATT CCG GGT GCG ACC ATG CAA GCA CCG ACC ACC GCG GCG AGC GCG GGT ACC TCG  
 trp cys ile arg ala ala ser met his ala pro thr ser ala ala ser ala ala ser cys  
 122/41  
 TCG CCG ACT GTC CAT CCA TCC CCG CAA CAT GTC TAA GAA TAT TAA GCG TAC TAC CG  
 leu pro thr val pro pro cys arg his his val arg his ala ala ala ala ala thr arg  
 182/61  
 TAT  
 his  
 242/81  
 GAT CAA TAT TAA AAT TAA CCG TTT TTA GCG GCG CCA TAA TAT TAT TAT TAT TAT TAT TAT  
 his gln asp arg asp asp arg leu pro ala gly pro asp lys cys val leu ala gly pro  
 302/101  
 CCG TAT GTC TAT TAA AAT TAA GCG TTT TTA GCG GCG CCA TAA TAT TAT TAT TAT TAT TAT TAT  
 his  
 362/121  
 TAT  
 his  
 422/141  
 TAT  
 his  
 482/161

33/11

FIGURE 6C

[illegible]

21/185

32/11  
 CGT CGG CAA CTT GGC CGC TGA GGT CGG CTT GAT CCC TGG GCC GAG GCG GGT CAG CCA ATA  
 arg arg gln leu gly arg OPA gly arg leu asp pro trp ala glu ala gly gln pro ile  
 62/21  
 GCG GCT CCA TCG GCT TTG CTG GTA GCG GTT CGG CGG GAA GCT AGC GGC GAC GTT GTC GGT  
 ala ala pro ser ala leu leu val ala val arg arg glu ala ser gly asp val val gly  
 122/41  
 GGC CGG TGA TAT ATT GGG TCA GAC GGG TAT GGC GGC GGC TGA GGT GAT CTG CGA CAC GCC  
 gly arg OPA tyr ile gly ser asp gly tyr gly gly gly OPA gly asp leu arg his ala  
 182/61  
 GCC GCG GTG CTC GAG CCA GGC TTA CGA CCA GGG AAT TTC GAA AAT GTT ATT CAG AAC ATC  
 ala ala val leu glu pro gly leu arg pro gly asn phe glu asn val ile gln asn ile  
 242/81  
 TTG TAT CTC TTC TCC CTG CCA CCC CCG AGG TGT AGT GTT TTC GAG TAC CGG CAG ATC CCA  
 leu tyr leu phe ser val pro pro pro arg cys ser val phe glu tyr arg gln ile pro  
 302/101  
 GGT TCA CCA GGT CTC ACC AGA TC  
 gly ser pro gly leu thr arg

SEQ ID N° 7B

FIGURE 7B

33/11  
 GTC GGC AAC TTS CCC GCT GAG GTC GGC TTG ATC CCT GCG CTC AGG CGS GTC AGC CAA TAG  
 val gly asn leu ala ala glu val gly leu ile pro gly pro arg arg val ser gln AMB  
 63/21  
 GGG CTC CAT CGG CTT TGC TGG TAG CGG TTC GGC GGC AAG CTA GCG GCG ACC TGG TGG GTC  
 arg leu his arg leu cys trp AMB arg phe gly gly lys leu ala ala thr leu ser val  
 123/41  
 GGC GGT GAT ATA TTS GGT CAG AGG GGT ATG GCG GCG GTT GAG GTC ATT TGT GAG AGG GTC  
 ala gly asp ile leu gly gln thr gly met ala ala ala ala val ile cys asp thr pro  
 183/61  
 CCG CGG TGG TGG AGG CAG GGT TAT GAT GAT GAT GAT ATT TTT AAA AAT TTA TTT AAA AAT TTT  
 pro arg cys ser ser ala ala thr arg ala ala ala ala ala ala ala ala ala ala ala  
 243/81  
 TGT ATG TGT TGT TGT TGT TGT TGT TGT TGT TGT TGT TGT TGT TGT TGT TGT TGT TGT TGT  
 cys ile leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu leu  
 303/101  
 GGT CAG CAG GGT TTA TTA GAT  
 val his gln val ser pro asp

SEQ ID N° 7C

FIGURE 7C

22/185

31/11  
 CTT TGC GTG ATG TCC AAT GGC GAA AAC GAC GCC TTG TCA TCG CAA TCG TCA GCA CCG GCC  
 leu cys val met ser asn gly glu asn asp ala leu ser ser gln ser ser ala pro ala  
 61/21  
 TAG TTT TCG CGA TGA CGC TCG TTC TGA CCG GAC TTG TGA ACG GGT TTC GGG TCG AGG CCG  
 AMB phe ser arg OPA arg ser phe OPA pro asp leu OPA thr gly phe gly ser arg pro  
 121/41  
 AGC GAA CCG TCG ATT CCA TGG GTG TCG ACG CAT TCG TGG TCA AGG CCG GCG CCG CAG GAC  
 ser glu pro ser ile pro trp val ser thr his ser trp ser arg pro ala arg gln asp  
 181/61  
 CGT TCC TGG GTT CGA CAC CAT TCG CCC AAA TCG ACC TGC CCC AGC TTG CTC GTG CCG CTG  
 arg ser trp val arg his his ser pro lys ser thr cys pro arg leu leu val arg leu  
 241/81  
 GCG TCT TGG CTG CCG CCC CAC TAG CGA CTG CCG CGT CGA CGA TCC GGC AGG GCA CGT CAG  
 ala ser trp leu pro pro his AMB arg leu arg arg arg arg ser gly arg ala arg gln  
 301/101  
 CGC GAA ACG TCA CCG CGT TCG GGG CAC CAG AGC ACG GAC CCG GCA TGC CCG GGG TCT CCG  
 arg glu thr ser pro arg ser gly his gln ser thr asp pro ala cys arg gly ser arg  
 361/121  
 ACG GTC GCG CCG CAT CGA CCG CCG ACG AGC TCG CCG TGT CGA GCA CCG TGG GCG GAA ACC  
 thr val gly arg his arg arg arg thr arg ser arg cys arg ala arg trp ala glu thr  
 421/141  
 TCG GCG ACG ATC  
 ser ala thr ile

SEQ ID N° 8A

FIGURE 8A

32/11  
 TTT GCG TGA TGT CCA ATG GCG AAA ACG ACG CCT TGT CAT CCG AAT TGT CAG CAC CCG GTT  
 pro ala OPA cys pro met ala lys thr thr pro cys ala arg asp arg gln his arg pro  
 62/21  
 AGT TTT CCG GAT GAC GGT CCG TTT GAC CCG ACT TGT GAA CCG GTT TCG GGT CGA GCG CGA  
 ser phe arg asp asp ala arg ser asp arg thr cys glu arg val ser gly arg gly arg  
 122/41  
 GCG AAC GGT CGA TTC CAT GAT GAT TTT CGA CCG ATT GGT GGT TAA GAT GAT TGT GCG ACG ACG  
 ala asp arg arg phe his gly cys arg arg ile arg gly ala gly arg arg gly arg thr  
 182/61  
 TTT GGT GAT GGT TCA AAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr  
 242/81  
 GAT GGT GAT TTT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 arg leu gly cys arg pro thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
 302/101  
 GCG AAA GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ala lys thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
 362/121  
 GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr

SEQ ID N° 8B

SEQUENCE ALIGNMENT FIGURE 8A

23/185

33/11  
 TTG CGT GAT GTC CAA TGG CGA AAA CGA CGC CTT GTC ATC GCA ATC GTC AGC ACC GGC CTA  
 leu arg asp val gln trp arg lys arg arg leu val ile ala ile val ser thr gly leu  
 63/21  
 93/31  
 GTT TTC GCG ATG ACG CTC GTT CTG ACC GGA CTT GTG AAC GGG TTT CGG GTC GAG GCC GAG  
 val phe ala met thr leu val leu thr gly leu val asn gly phe arg val glu ala glu  
 123/41  
 153/51  
 CGA ACC GTC GAT TCC ATG GGT GTC GAC GCA TTC GTG GTC AAG GCC GGC GCG GCA CGA CCG  
 arg thr val asp ser met gly val asp ala phe val val lys ala gly ala ala gly pro  
 183/61  
 213/71  
 TTC CTG GGT TCG ACA CCA TTC GCG CAA ATC GAC CTG CCC CAG GTT GGT CGT GCG CCT GGC  
 phe leu gly ser thr pro phe ala gln ile asp leu pro gln val ala arg ala pro gly  
 243/81  
 273/91  
 GTC TTG GCT GCG GCG CCA CTA GCG ACT GCG CCG TCG ACG ATC CCG CAG GCG ACG TCA GCG  
 val leu ala ala ala pro leu ala thr ala pro ser thr ile arg gln gly thr ser ala  
 303/101  
 333/111  
 CGA AAC GTC ACC GCG TTC GGG GCA CCA GAG CAC GGA CCG GGC ATG CCG CCG GTC TCG GAC  
 arg asn val thr ala phe gly ala pro glu his gly pro gly met pro arg val ser asp  
 363/121  
 393/131  
 GST CGG GCG CCA TCG ACG CCG GAC GAG GTC CCG GTG TCG ACC ACG CTG GGC CGA AAC CTC  
 gly arg ala pro ser thr pro asp glu val ala val ser ser thr leu gly arg asn leu  
 423/141  
 GGC GAC GAT C  
 gly asp asp

SEQ ID N° 8C

FIGURE 8C

partie de la séquence nucléotidique de seq8A

1/1  
 31/11  
 CAG GTT GGT GGT GCG CCT GCG GTC TTG GGT GCG GCG CCA CTA GCG ACT GCG CCG TCG ACC  
 gln val ala arg ala pro gly val leu ala ala ala pro leu ala thr ala pro thr  
 61/21  
 91/31  
 ATT CCG CAG CAG AAT TTA CCG CCA AAT ATC ACG GGT TTC GAT GAT CCA GAT CAG CCA GAT  
 ile arg ala gly thr ser ala arg ala val thr phe phe gly phe ala phe thr  
 121/41  
 151/51  
 GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 gly met pro pro pro pro pro pro pro pro pro pro pro pro pro pro pro pro  
 181/61  
 AAC AAT CTA AAT CCA AAT CTA GCG GAT GAT C  
 asn thr leu arg arg ala leu gly arg asp

SEQ ID N° 8A'

24/185

1/1 31/11  
 AGG TTG CTC GTG CGC CTG GCG TCT TGG CTG CCG CCC CAG TAG CGA CTG CGC CGT CGA CGA  
 arg leu leu val arg leu ala ser trp leu pro pro his AMB arg leu arg arg arg arg  
 61/21 91/31  
 TCC GGC AGG GCA CGT CAG CGC GAA ACG TCA CCG CGT TCG GGG CAC CAG AGC ACG GAC CCG  
 ser gly arg ala arg gln arg glu thr ser pro arg ser gly his gln ser thr asp pro  
 121/41 151/51  
 GCA TGC CGC GGG TCT CGG ACG GTC GGG CGC CAT CGA CGC CGG ACG AGG TCG CGG TCT CGA  
 ala cys arg gly ser arg thr val gly arg his arg arg arg thr arg ser arg cys arg  
 181/61  
 GCA CGC TGG GCC GAA ACC TCG GCG ACG ATC  
 ala arg trp ala glu thr ser ala thr ile

SEQ ID N° 8B'

FIGURE 8B'

Seq8C

1/1 31/11  
 CCA GGT TGC TCG TGC GCC TGG CGT CTT GGC TGG GCG GCG AGT AGC GAC TGC GCG GTC GAC  
 pro gly cys ser cys ala trp arg leu gly tyr arg pro thr ser asp cys ala val asp  
 61/21 91/31  
 GAT CCG GCA GCG CAG CTC AGC GCG AAA CTT CAT TCT CTT CAG GGT AGC AGA GCA CCG ACC  
 asp pro ala gly his val ser ala lys arg his arg val arg gly thr arg ala arg thr  
 121/41 151/51  
 CCG CAT GCG GCG GGT CTC GGA CCG TCG GGT CTT AGT CAT TCT GAA GAA AGT CCG GGT CTT  
 arg his ala ala gly leu gly arg ser gly his arg arg ala gly arg gly arg gly val  
 181/61  
 GAG CAG GGT GCG CCG AAA CTT CAG GAA GAA T  
 ala his ala gly pr gly pro thr thr thr

SEQ ID N° 8C'

FIGURE 8C'



séquence Rv2563 prédite par Cole et al. (Nature 393:537-544) et contenant seq8A'

© 2004 Blackwell Publishing Ltd, *Journal of Internal Medicine* 255: 111–119

26/185

ORF prédite par\_Cole et al. (Nature 393:537-544) et contenant Rv2563

1/1  
tag gtt tca aga agg cct gtg cag gtt tcc gca gcc tgg gcc gcg gcg cca ccg aag agc  
AMB val ser arg arg pro val gln val ser ala ala trp ala ala ala pro pro lys ser  
61/21  
ccg ccg aaa tgg gct aat ccg gtt cgc ttg gct cga tgc ccg atg atc tgc acc gcc acc  
pro pro lys trp ala asn arg val arg leu ala arg ser pro met ile ser thr ala thr  
121/41  
acc gac ccc ctc acc tgc gtc gaa cct ccg cga acc aac gcg gca acc cca gcc cat gat  
thr asp pro leu thr ser val glu pro arg arg thr asn ala ala thr pro ala his asp  
181/61  
cat ttg att ggg tcc acg gaa gca ggt agc ttc cgt cgc atg ctt ttt gcg gct ttg cgt  
his leu ile gly ser thr glu ala gly ser phe arg arg met leu phe ala ala leu arg  
241/81  
gat gtc caa tgg cga aaa cga cgc ctt gtc atc gca atc gtc agc acc gcc cta gtt ttc  
asp val gln trp arg lys arg arg leu val ile ala ile val ser thr gly leu val phe  
301/101  
gcg atg acg ctc gtt ctg acc gga ctt gtg aac gcg ttt cgg gtc gag gcc gag cga acc  
ala met thr leu val leu thr gly leu val asn gly phe arg val glu ala glu arg thr  
361/121  
gtc gat tcc atg ggt gtc gac gca ttc gtg gtc aag gcc gcc gcg gca gga ccg ttc ctg  
val asp ser met gly val asp ala phe val val lys ala gly ala ala gly pro phe leu  
421/141  
ggt tgc aca cca ttc gcc caa atc gac ctg ccc cag gtt gat cgt gcg cct gcc gtc ttg  
gly ser thr pro phe ala gln ile asp leu pro gln val ala arg ala pro gly val leu  
481/161  
gct gcc gcc cca cta gcg act gcg ccg tgc acc atc ccg cag gcc acc tca gcg cga aac  
ala ala ala pro leu ala thr ala pro ser thr ile arg gln gly thr ser ala arg asn  
541/181  
gtc acc gcg ttc ggg gca cca gag cac gga ccc gcc atg ccg ccg gtc tgc gac ggt ccg  
val thr ala phe gly ala pro glu his gly pro gly met pro arg val ser asp gly arg  
601/201  
gcg cca tgc acg ccg gac gag gtc gcg gtg tgc acc ccg ctg gcc cga aac ctc gcc gac  
ala pro ser thr pro asp glu val ala val ser ser thr leu gly arg asn leu gly asp  
661/221  
gat ctg caa ctg ggt gcg ccg act ttg ccg atc gtc gcc atc gtg ccc gag tca acc gcc  
asp leu gln val gly ala arg thr leu arg ile val gly ile val pro glu ser thr ala  
721/241  
ctg gca aag att ccc aac atc ttc ctg acc acc gaa gcc cta cag cag ttg gca tac aac  
leu ala lys ile pro asn ile phe leu thr thr glu gly leu gln gln leu ala tyr asn  
781/261  
gga cag ccg ala arg agt tgc arg ggg atc gac gcc atg ccc cca cag ctg ccg gcc gcc  
gly gln pro thr ile ser ser ile gly ile asp gly met pro arg gln leu pro asp gly  
841/281  
tat cag acc ctg aat cca gcc gat gtt gtc agc gat ctg ala cgt ccg ttc aag ctg gcc  
tyr gln thr val arg arg ala asp ala val ser asp leu met arg pro leu lys val ala  
901/301  
ctg gat gcc atc acc att gta ggt gtc ttc ctg tgc atc gtc gtc gtc gtc gtc gtc gtc  
val asp ala ile thr val val ala val leu leu trp ile val ala ala leu ile val gly  
961/321  
tat ctg gtc ttc atc tat gac ttc cag cga ctg cct gcc ttc ggc atg ttc aat ggc atg  
ser val val tyr leu ser ala leu leu arg leu arg asp phe ala val phe lys ala ile  
1021/341  
gac gtc gcc ttc gcc ttc att ctg ttc ggg cta gcc atc gcc gtc atc gcc gtc gcc  
gly val pro thr arg ser ile leu ala gly leu ala leu ala ala val val val ala leu  
1081/361  
tat gtc gtc ttc gcc ttc atc ttc gtc ttc gtc ttc gtc ttc gtc ttc gtc ttc gtc ttc  
ser val val tyr leu ser ala leu leu arg leu arg asp phe ala val phe lys ala ile  
1141/381  
gac gtc gcc ttc gcc ttc att ctg ttc ggg cta gcc atc gcc gtc atc gcc gtc gcc  
gly val pro thr arg ser ile leu ala gly leu ala leu ala ala val val val ala leu  
1201/401  
tat gtc gtc ttc gcc ttc atc ttc gtc ttc gtc ttc gtc ttc gtc ttc gtc ttc gtc ttc  
ser val val tyr leu ser ala leu leu arg leu arg asp phe ala val phe lys ala ile  
1261/421

27/185

séquence de Rv0072 prédite par par Cole et al. (Nature 393:537-544) et  
présentant plus de 77% de similarité avec Seq8D'

1/1 31/11  
atg ctc ttc gcg gcc ctg cgt gac atg caa tgg aga aag cgc cgc ctg gtc atc acg atc  
Met leu phe ala ala leu arg asp met gln trp arg lys arg arg leu val ile thr ile  
61/21 91/31  
atc agc acc ggg ctg atc ttc ggg atg acg ctt gtt ttg acc gga ctc gcg aac ggc ttc  
ile ser thr gly leu ile phe gly met thr leu val leu thr gly leu ala asn gly phe  
121/41 151/51  
cgg gtg gag gcc cgg cac acc gtc gat tcc atg ggt gtc gat gta ttc gtc gtc aga tcc  
arg val glu ala arg his thr val asp ser met gly val asp val phe val val arg ser  
181/61 211/71  
ggc gct gct gga cct ttt ctg ggt tca ata ccg ttt ccc gat gtt gac ctg gcc cga gtg  
gly ala ala gly pro phe leu gly ser ile pro phe pro asp val asp leu ala arg val  
241/81 271/91  
ccc gct gaa ccc ggt gtc atg gcc gcg gcc ccg ttg ggc agc gtg ggg acg atc atg aaa  
ala ala glu pro gly val met ala ala ala pro leu gly ser val gly thr ile met lys  
301/101 331/111  
gaa ggc acg tcg acg cga aac gtc acg gtc ttc ggc gcg ccc gag cac gga cct ggc atg  
glu gly thr ser thr arg asn val thr val phe gly ala pro glu his gly pro gly met  
361/121 391/131  
cca cgg gtc tca gag ggt cgg tca ccg tcg aaa ccg gac gaa gtc gcg gca tcg agc acg  
pro arg val ser glu gly arg ser pro ser lys pro asp glu val ala ala ser ser thr  
421/141 451/151  
atg ggc cga cac ctc ggt gac act gtc gag gtc ggc gcg cgc aga ttg cgg gtc gtt ggc  
met gly arg his leu gly asp thr val glu val gly ala arg arg leu arg val val gly  
481/161 511/171  
att gtg ccg aat tcc acc gcg ctg gcc aag atc ccc aat gtc ttc ctc acg acc gag ggc  
ile val pro asn ser thr ala leu ala lys ile pro asn val phe leu thr thr glu gly  
541/181 571/191  
tta cag aaa ttg gcg tac aac ggg cag ccg aat atc acg tcc atc ggg atc ata ggt atg  
leu gln lys leu ala tyr asn gly gln pro asn ile thr ser ile gly ile ile gly met  
601/201 631/211  
ccc cga cag ctg ccg gag ggt tac cag act ttc gat cgg gtg ggc gct gtc aat gat ttg  
pro arg gln leu pro glu gly tyr gln thr phe asp arg val gly ala val asn asp leu  
661/221 691/231  
gtg cgc cca ttg aag gtc gca gtg aat tcg atc tgg atc gtg gct gtc ttg ctg tga att  
val arg pro leu lys val ala val asn ser ile ser ile val ala val leu leu trp ile  
721/241 751/251  
gtg ccc gtg cgg atc gtc gln tcg atg atc ttc att tgg gat ttc acc ggt cta trp ca  
val ala val leu ile val gly ser val val tyr leu ser ala leu gln arg leu arg asp  
781/261 811/271  
ttc gct ttc ttc aac gct atc ggc aag ttc aac ggc ttc atc atc ttc ggc atc ttc ttc  
phe ala val phe lys ala ile gly thr pro thr arg ser ile met ala gly leu ala leu  
841/281 871/291  
cag gct ctg gtc atc ggc ttg att gag cgg atc ttc gtc atc atc atc atc atc atc atc  
ala ala leu val ile ala leu leu ala ala val val gly val val leu ala gln val leu  
901/301 931/311  
tca cca gta ttc gca ata ttc ggc gta ttc ttc ggc gta ttc ttc ggc gta ttc ttc ggc  
gta pro leu phe pro met ala val ala val ala val gly ala tyr leu ala leu pro gln  
961/321 991/331  
tca cca gta ttc gca ata ttc ggc gta ttc ttc ggc gta ttc ttc ggc gta ttc ttc ggc  
gta pro leu phe pro met ala val ala val ala val gly ala tyr leu ala leu pro gln

SEQ. ID. NO. 41

SEQUENCE ALIGNMENT

28/185

Seq8H : ORF prédite par Cole et al. (Nature 393:537-544) et contenant seq8G  
1/1 31/11  
tag cct ctg gga atg ctc ttc gcg gcc ctg cgt gac atg caa tgg aga aag cgc cgc ctg  
AMB pro leu gly met leu phe ala ala leu arg asp met gln trp arg lys arg arg leu  
61/21 91/31  
gtc atc acg atc atc agc acc ggg ctg atc ttc ggg atg acg ctt gtt ttg acc gga ctc  
val ile thr ile ile ser thr gly leu ile phe gly met thr leu val leu thr gly leu  
121/41 151/51  
gcg aac ggc ttc cgg gtg gag gcc cgg cac acc gtc gat tcc atg ggt gtc gat gta ttc  
ala asn gly phe arg val glu ala arg his thr val asp ser met gly val asp val phe  
181/61 211/71  
gtc gtc aga tcc ggc gct gct gga cct ttt ctg ggt tca ata cgg ttt ccc gat gtt gac  
val val arg ser gly ala ala gly pro phe leu gly ser ile pro phe pro asp val asp  
241/81 271/91  
ctg gcc cga gtg gcc gct gaa ccc ggt gtc atg gcc gcg gcc cgg ttg ggc agc gtg ggg  
leu ala arg val ala ala glu pro gly val met ala ala ala pro leu gly ser val gly  
301/101 331/111  
acg atc atg aaa gaa ggc acg tcc acg cga aac gtc acg gtr ttc ggc ggc ccc gag cac  
thr ile met lys glu gly thr ser thr arg asn val thr val phe gly ala pro glu his  
361/121 391/131  
gga cct ggc atg cca cgg gtc tca gag ggt cgg tca cgg tcc aaa cgg gac gaa gtc gcc  
gly pro gly met pro arg val ser glu gly arg ser pro ser lys pro asp glu val ala  
421/141 451/151  
gca tcc agc acg atg ggc cga cac ctc ggt gac act gtc gag gtc ggc gcg cgc aga ttg  
ala ser ser thr met gly arg his leu gly asp thr val glu val gly ala arg arg leu  
481/161 511/171  
cgg gtc gtt ggc att gtg ccg aat tcc acc gcg ctg gcc aag ata ccc aat gtc ttc ctc  
arg val val gly ile val pro asn ser thr ala leu ala lys ile pro asn val phe leu  
541/181 571/191  
acg acc gag ggc tta cag aaa ttg gag tac aac ggc cag cgg aat ata acc tcc ata ggg  
thr thr glu gly leu gln lys leu ala tyr asn gly gln pro asn ile thr ser ile gly  
601/201 631/211  
atc ata ggt atg ccc cga cag ctg ccc cag ggt tta cag aat tta ggt cgg gtg gcc gct  
ile ile gly met pro arg gln leu pro glu gly tyr gln thr phe asp arg val gly ala  
661/221 691/231  
gtc aat gat ttg gtg gcc cca ttg aag gtc gtr arg gtr tta ata tgg ata ggg gct gtt  
val asn asp leu val arg pro leu lys val ala val asn ser ile ser ile val ala val  
721/241 751/251  
ctg ctg tta att atg tgg tta cta atc gtc tta tta tta tta tta tta tta tta tta tta  
leu leu trp ile val ala val leu ile val gly ser val val tyr leu ser ala leu ala  
781/261 811/271  
ctg ata gtr gtr tta tta gtr tta att gtr tta tta tta tta tta tta tta tta tta tta  
arg leu arg asp phe ala val phe lys ala ile gly thr pro thr arg ser thr ser ala  
841/281 871/291  
ggg ctg tta tta ggt gtr ttg tta att gct tta att tta gtr tta tta gtr gtr gtr gtr  
gly leu ala leu gln ala leu val ala ala leu leu ala ala val val gly val val leu  
901/301 931/311  
ctg gtr tta tta gtr tta tta tta tta tta tta tta tta tta tta tta tta tta tta  
ala gln val ala gln gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr  
961/321 991/331  
gtr tta gtr tta gtr tta tta tta tta tta tta tta tta tta tta tta tta tta tta  
ala gln val ala gln gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr  
1021/341 1051/351  
gtr tta gtr tta gtr tta tta tta tta tta tta tta tta tta tta tta tta tta tta  
ala gln val ala gln gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr gtr  
1081/361 1111/371

Seq 8H : ORF

Seq 8H : ORF prédite par Cole et al. (Nature 393:537-544) et contenant seq8G

29/185

31/11  
 CGA GGC CGA GCG AAC CGT CGA TTC CAT GGG TGT CGA CGC ATT CGT GGT CAA GGC CGG CGC  
 arg gly arg ala asn arg arg phe his gly cys arg arg ile arg gly gln gly arg arg  
 61/21  
 GGC AGG ACC GTT CCT GGG TTC GAC ACC ATT CGC CCA AAT CGA CCT GCC CCA GGT TGC TCG  
 gly arg thr val pro gly phe asp thr ile arg pro asn arg pro ala pro gly cys ser  
 121/41  
 TGC GCC TGG CGT CTT GGC TGC CGC CCC ACT AGC GAC TGC GCC CTC GAC GAT CCG GCA GGG  
 cys ala trp arg leu gly cys arg pro thr ser asp cys ala val asp asp pro ala gly  
 181/61  
 CAC GTC ACC GCG AAA CGT CAC CGC GTT CCG GGC ACC AGA GCA CGG ACC CGG CAT GCC GCG  
 his val ser ala lys arg his arg val arg gly thr arg ala arg thr arg his ala ala  
 241/81  
 GGT CTC GGA CGG TCG GGC GCC ATC GAC GCC GGA CGA GGT CGC GGT GTC GAG CAC GGT GGC  
 gly leu gly arg ser gly ala ile asp ala gly arg gly arg gly val glu his ala gly  
 301/101  
 CCG AAA CCT CGG CGA CGA TC  
 pro lys pro arg arg arg

SEQ ID N° 9A

FIGURE 9A

32/11  
 GAG GCC GAG CGA ACC GTC GAT TCC ATG GGT GTT GAC GGA TTT GTS GTC AAG GCC GGG GCG  
 glu ala glu arg thr val asp ser met gly val asp ala phe val val lys ala gly ala  
 62/21  
 GCA GGA CCG TTC CTG GGT TCG ACA CCA TTC GGT CAA ATC GAG TTT GCG GAG GTT GGT GGT  
 ala gly pro phe leu gly ser thr pro phe ala gln ile asp leu pro ala val ala arg  
 122/41  
 GCG CCT GCG GTC TTG GGT GCC GCG CCA CTA GCG ACT GCG GCG TTS AGG ATC CCG CAG GGT  
 ala pro gly val leu ala ala ala pro leu ala thr ala pro ser thr ile arg gln gly  
 182/61  
 ACG TCA GCG GGA AAT GTC ACT GCG TTG GCG GGA CCA GAA AAT GGA GAA AAT GGA GAA AAT GGA  
 thr ser ala arg asn val thr ala phe gly ala pro ala gln gly pro gly met pro arg  
 242/81  
 GGT TTT GAT GAT GGT GGT GGA TTT AAT GAA GAA GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 val ser gln gly thr asp pro thr thr pro pro pro pro pro pro pro pro pro pro pro pro  
 302/101  
 GGA AAA TTT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 arg val gln gly asp arg

SEQ ID N° 9B

30/185

33/11  
 AGG CCG AGC GAA CCG TCG ATT CCA TGG GTG TCG ACG CAT TCG TGG TCA AGG CCG GCG CGG  
 arg pro ser glu pro ser ile pro trp val ser thr his ser trp ser arg pro ala arg  
 63/21  
 CAG GAC CGT TCC TGG GTT CGA CAC CAT TCG CCC AAA TCG ACC TGC CCC AGG TTG CTC GTG  
 gln asp arg ser trp val arg his his ser pro lys ser thr cys pro arg leu leu val  
 123/41  
 CGC CTG GCG TCT TGG CTG CCG CCC CAC TAG CGA CTG CGC CGT CGA CGA TCC GGC AGG GCA  
 arg leu ala ser trp leu pro pro his AMB arg leu arg arg arg arg ser gly arg ala  
 183/61  
 CGT CAG CGC GAA ACG TCA CCG CGT TCG GGG CAC CAG AGC ACG GAC CCG GCA TGC CGC GGG  
 arg gln arg glu thr ser pro arg ser gly his gln ser thr asp pro ala cys arg gly  
 243/81  
 TCT CGG ACG GTC GGG CGC CAT CGA CGC CGG ACG AGG TCG CGG TGT CGA GCA CGC TGC GCC  
 ser arg thr val gly arg his arg arg arg thr arg ser arg cys arg ala arg trp ala  
 303/101  
 GAA ACC TCG GCG ACG ATC  
 glu thr ser ala thr ile

SEQ ID N° 9C

FIGURE 9C

31/11  
 TTA ACG ACT CAG ACG GAA ACG CTT GAA CCG CGA GGT CGG TCG GGA CAC CAA TTT GAC TCG  
 leu thr thr gln thr glu thr leu glu pro arg gly arg ser gly his gln phe asp ser  
 61/21  
 GGT CTT TGG CAA TTG AAG GTG AGC TCG GAG CAG CGG GGT GAC CGC ATC GTT GCG CTT GCG  
 ala leu trp gln leu lys val ser cys glu gln pro gly asp arg ile val gly leu ala  
 121/41  
 ATC AAT CGC CGC CTC GCG GAC GTA GAT AAT CAG CTC ATT GTT GCG ACC GAC CTC GAC CAG  
 ile asn arg arg leu ala asp val asp asn gln thr val gly thr asp leu asp gln  
 181/61  
 GGG TCG TTT GTG ACT GGT GGG CTT GAG GCG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG  
 gly ser phe val thr ala gly leu asp ala asp asp ala asp val gly his arg leu arg  
 241/81  
 TTA CGT TTT TGA CTT GGT GGT GGT TGG GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT  
 leu pro phe TTA pro gly ala ala thr ala pro thr thr TTA gly thr thr thr thr thr  
 301/101  
 GGT CGA CGG CGA CTT CGG TCG GCG GCA GTA TGT GAG CAT GTG GAG ATG ATT GCA GGT AAT  
 gly pro pro pro pro arg ser pro ala val cys gln leu val ile pro thr pro arg ser  
 361/121  
 TTT GGT GGT AAT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT  
 thr val thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
 421/141  
 TTT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT GGT  
 thr

SEQUENCE 1A

31/185

32/11  
TAA CGA CTC AGA CGG AAA CGC TTG AAC CGC GAG GTC GCT CCG GAC ACC AAT TTG ACT CGG  
OCH arg leu arg arg lys arg leu asn arg glu val ala pro asp thr asn leu thr arg  
62/21  
CTC TTT GGC AAT TGA AGG TGA GCT GCG AGC AGC CGG GTG ACC GCA TCG TTG GCC TTG CCA  
leu phe gly asn OPA arg OPA ala ala ser ser arg val thr ala ser leu ala leu pro  
122/41  
TCA ATC GCC GGC TCG CGG ACG TAG ATA ATC AGC TCA CCG TTG GGA CCG ACC TCG ACC AGG  
ser ile ala gly ser arg thr AMB ile ile ser ser pro leu gly pro thr ser thr arg  
182/61  
GGT CCT TTG TGA CTG CCG GGC TTG ACG CGG ACG ACC ACA GAG TCG GTC ATC GCC TAA GGU  
gly pro leu OPA leu pro gly leu thr arg thr thr thr glu ser val ile ala OCH gly  
242/81  
TAC CGT TCT GAC CTG GGG CTG CGT GGG CGC CGA CGA CGT GAG GCA CGT CAT GTC TCA GCG  
tyr arg ser asp leu gly leu arg gly arg arg arg arg glu ala arg his val ser ala  
302/101  
GCG CAC CGC GAC CTC GGT CGC CGG CAG TAT GTG AGC ATG TGC AGA TGA CTC CAC GCA GCC  
ala his arg his leu gly arg arg gln tyr val ser met cys arg OPA leu his ala ala  
362/121  
TTG TTC GCA TCG TTG GTG TCG TGG TTG CGA CGA CCT TCG CGC TGG TGA GCG CAC CGG CGG  
leu phe ala ser leu val ser trp leu arg arg pro trp arg trp OPA ala his pro pro  
422/141  
GGG GTC GTG CCG CGC ATG CGG ATC  
ala val val pro arg met arg ile

SEQ ID N° 10B

FIGURE 10B

33/11  
AAG GAC TCA GAC GGA AAC GCT TGA ACG GCG AGG TCG CTC CGG ACA CGA ATT TGA CTC GGT  
asn asp ser asp gly asn ala OPA thr ala arg ser leu arg thr pro ile OPA leu gly  
68/21  
TTT TTG GCA ATT GAA GGT GAG CTG CGA GCA GCG GGG TGA CGA CAT CAT TGG GGT TGG CAT  
ser leu ala ile glu gly glu leu arg ala ala gly OPA pro his arg trp pro cys his  
128/41  
AAA TCG GCG GGT TGT GGA CGT AGA TAA TGA ATT CAG CAT TGG GAT GGA GAT GGA GGA GGA  
asn ser pro ala ala arg arg arg OCH ser ala his arg trp asp arg pro arg pro gly  
188/61  
GAT CTT TAA GAT TAT GAT TGT TGA CA GAA TGA TGA AAG GAT TAT TAT TAT TAT TAT  
asp leu tyr arg cys ala ala OPA arg arg arg pro ala ser his pro arg arg arg  
248/81  
AAG GTT CTT AAT TTA GAT TGT GTG GAT TGA GAT GAT GAT AAT GAT GAT AAT TAT TAT GAT  
thr val leu thr thr gly cys val gly ala asp asp val arg his val met ser thr arg  
308/101  
AGG AAG GAT AAT TAT GAT TGT GAT AAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT  
arg arg asp asp asp asp asp asp asp asp asp asp asp asp asp asp asp asp asp  
368/121  
GAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT  
asp thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
428/141  
GAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT  
asp thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

Seq. ID N° 11

Sequence alignment with SEQ ID N° 1

32/185

31/11  
 CCC GAA GAG GTC CCC CGT TTT GTT AAT TTT TAA AAA ATT TGT GTC ACA AAC CGG GGT ACC  
 pro glu glu val pro arg phe val asp phe OCH lys ile cys val thr lys arg gly thr  
 61/21  
 AAG GCA TAA AAC CTA GTA CCT GGG GCG GCG GAT TCA ACG AAA ACC GAG TGG GGG TAG TCA  
 lys ala OCH asn leu val pro gly ala ala asp ser thr lys thr glu trp gly AMB ser  
 121/41  
 GGG GCG TGC ATT CCG ACG ACC CTG TAC GAC CCG CTG GTG GCA ACG CCG ATG AGT GCG CCG  
 gly ala cys ile pro thr thr leu tyr asp pro leu val ala thr pro met ser ala pro  
 181/61  
 ACG AAG GCC GAG CGA CCG GCT GCC GGC GCT GAC CGC CGC GGA AGC CGC CGA GTG CAT GGT  
 thr lys ala glu arg arg ala ala gly ala asp arg arg gly ser arg arg val asp gly  
 241/81  
 CAC CAC CGC CCG CAC CCG ACC GGT ACG GAT CGC GCC TCG GGT TAC CGT CGC CGT CAA CGC  
 his his arg pro his pro thr gly thr asp arg ala ser gly tyr arg arg arg gln arg  
 301/101  
 GCT GGA CAG CAT CCG TCC CCC CTC GGT CAA TGC ACT CAC AAA GCG CCG CAA CGA ACA GCT  
 ala gly gln his arg ser pro leu gly gln sys thr his ala ala pro gln arg thr ala  
 361/121  
 CAA CCC TTS AAC CCG GTC CCG GCC TGC CGA CCC TCG GCG GCG GCG GTG CCG CTA CGT GAT  
 gln pro leu asn arg val pro ala sys arg pro ser ala ala gly val pro leu arg asp  
 421/141  
 AGA CAC AGC GCC ATG GAA ATC CTG TCC AGC CCG ATG CTA GTT CCG CCG CCG GAC TAT CAG  
 arg his arg ala met glu ile leu ala ser arg met leu leu arg pro ala asp tyr gln  
 481/161  
 CGS TCG CTC AGC TTC TAC CGT GAC CAG ATC  
 arg ser leu ser phe tyr arg asp gln ile

SEQ ID N° 11A

FIGURE 11A

32/11  
 CCG AAG AGG TCC CCC GTT TTS TTA ATT TTT AAA AAA TTT CTC TTA CAA AGC GGG GTA CCA  
 pro lys arg ser pro val leu leu ile phe lys lys phe val ser gln ser gly val pro  
 62/21  
 AGG CAT AAA ACC TAG TAC CTC GGG CCG CCG ATT CAA CAA AAA CCG ATT TGG GGT AGT CAG  
 arg his lys thr AMB tyr leu gly arg arg ile gln arg lys pro ser gly gly ser gln  
 122/41  
 GCG CGT CCA TTC CGA CGA CCG TGT ACG ACC CCG TCG TTA TAA CCA CCA TAA GTG CCG CGA  
 gly arg ala phe arg arg pro cys thr thr arg thr thr arg arg arg arg val arg arg  
 182/61  
 CAA AGG CCG AAG GAG GCG CTC CCG GCG CTC ACC CCG CCG CAA CCA CCA CCA TCG ATG TTC  
 gln arg phe arg asp gly leu pro ala leu thr arg arg arg ala ala thr thr met val  
 242/81  
 ACC AGC CCG GGT ACC CGA CCG GGA CCG ATG AAG CTC TCA TTA ATT CTC GGT CTC TTC ACC GGT  
 thr thr ala arg thr arg pro val arg ile ala pro arg val thr val ala val arg ala  
 302/101  
 CTG GAG AGC ATT GGT CAC GGT TCG GTC AAT GCA TTC AAG CAG GAT TTA AAT GAA GAA GAT  
 leu asp arg ile gly pro arg thr val gln ala leu arg gln arg arg arg gln gln leu  
 362/121  
 AAT CTC TTA ATG GAG CCG CCG CTC CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG  
 asn thr TAA ATG GAG CCG CCG CTC CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG  
 422/141  
 GAT CTC TTA ATG GAG CCG CCG CTC CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG CCG  
 482/161



33/185

33/11  
 CGA AGA GGT CCC CCG TTT TGT TAA TTT TTA AAA AAT TTG TGT CAC AAA GCG GGG TAC CAA  
 arg arg gly pro pro phe cys OCH phe leu lys asn leu cys his lys ala gly tyr gln  
 63/21  
 GGC ATA AAA CCT AGT ACC TGG GGC GGC GGA TTC AAC GAA AAC CGA GTG GGG GTA GTC AGG  
 gly ile lys pro ser thr trp gly gly gly phe asn glu asn arg val gly val val arg  
 123/41  
 GGC GTG CAT TCC GAC GAC CCT GTA CGA CCC GCT GGT GGC AAC GCC GAT GAG TGC GCC GAC  
 gly val his ser asp asp pro val arg pro ala gly gly asn ala asp glu cys ala asp  
 183/61  
 GAA GGC CGA GCG ACG GGC TGC CGG CGC TGA CCG CCG CCG AAG CCG CCG AGT GGA TGG TCA  
 glu gly arg ala thr gly cys arg arg OPA pro pro arg lys pro pro ser gly trp ser  
 243/81  
 CCA CCG CCC GCA CCC GAC CGG TAC GGA TCG CCG CTC GGG TTA CCG TCG CCG TCA ACG CCG  
 pro pro pro ala pro asp arg tyr gly ser arg leu gly leu pro ser pro ser thr arg  
 303/101  
 TGG ACA GCA TCG GTC CCC GCT GGG TCA ATG CAC TCA TGG AGC GCG GCA ACG AAC AGC TCA  
 trp thr ala ser val pro ala gly ser met his ser cys ser ala ala thr asn ser ser  
 363/121  
 ACC CTT GAA CCG GGT CCC GGC CTG CCG ACC CTC GGC CCG CCG CCG GCG GGT ACG TGA TAG  
 thr leu glu pro gly pro gly leu pro thr leu gly arg arg arg ala ala thr OPA AMB  
 423/141  
 ACA CAG GGC CAT GGA AAT CCT GGC CAG CCG GAT GCT ACT TCG GCG GGC GGA CTA TCA GCG  
 thr gln gly his gly asn pro gly gln pro asp ala thr ser ala gly gly leu ser ala  
 483/161  
 GTC GCT GAG CTT CTA CCG TGA CCA GAT C  
 val ala glu leu leu pro OPA pro asp

SEQ ID N° 11C

FIGURE 11C

partie de la séquence nucléotidique de SEQ  
 171  
 CGA AGA GGT CCC CCG TTT TGT TAA TTT TTA AAA AAT TTG TGT CAC AAA GCG GGG TAC CAA  
 arg arg gly pro pro phe cys OCH phe leu lys asn leu cys his lys ala gly tyr gln  
 61/21  
 GGC ATA AAA CCT AGT ACC TGG GGC GGC GGA TTC AAC GAA AAC CGA GTG GGG GTA GTC AGG  
 gly ile lys pro ser thr trp gly gly gly phe asn glu asn arg val gly val val arg  
 121/41  
 GGC GTG CAT TCC GAC GAC CCT GTA CGA CCC GCT GGT GGC AAC GCC GAT GAG TGC GCC GAC  
 gly val his ser asp asp pro val arg pro ala gly gly asn ala asp glu cys ala asp  
 181/61  
 GAA GGC CGA GCG ACG GGC TGC CGG CGC TGA CCG CCG CCG AAG CCG CCG AGT GGA TGG TCA  
 glu gly arg ala thr gly cys arg arg OPA pro pro arg lys pro pro ser gly trp ser  
 241/81  
 CCA CCG CCC GCA CCC GAC CGG TAC GGA TCG CCG CTC GGG TTA CCG TCG CCG TCA ACG CCG  
 pro pro pro ala pro asp arg tyr gly ser arg leu gly leu pro ser pro ser thr arg  
 301/101  
 TGG ACA GCA TCG GTC CCC GCT GGG TCA ATG CAC TCA TGG AGC GCG GCA ACG AAC AGC TCA  
 trp thr ala ser val pro ala gly ser met his ser cys ser ala ala thr asn ser ser  
 361/121  
 ACC CTT GAA CCG GGT CCC GGC CTG CCG ACC CTC GGC CCG CCG CCG GCG GGT ACG TGA TAG  
 thr leu glu pro gly pro gly leu pro thr leu gly arg arg arg ala ala thr OPA AMB  
 421/141  
 ACA CAG GGC CAT GGA AAT CCT GGC CAG CCG GAT GCT ACT TCG GCG GGC GGA CTA TCA GCG  
 thr gln gly his gly asn pro gly gln pro asp ala thr ser ala gly gly leu ser ala  
 481/161  
 GTC GCT GAG CTT CTA CCG TGA CCA GAT C  
 val ala glu leu leu pro OPA pro asp

1/1	31/11																		
GTC GCC GTC AAC GCG CTG GAC AGC ATC GGT	CCC CGC TGG GTC AAT GCA CTC ATG CAG CGC																		
val ala val asn ala leu asp ser ile gly	pro arg trp val asn ala leu met gln arg																		
61/21	91/31																		
CGC AAC GAA CAG CTC AAC CCT TGA ACC GGG	TCC CGG CCT GCC GAC CCT CGG CCG CCG GCG																		
arg asn glu gln leu asn pro OPA thr gly	ser arg pro ala asp pro arg pro pro ala																		
121/41	151/51																		
TGC CGC TAC GTG ATA GAC ACA GGG CCA TGG	AAA TCC TGG CCA GCC GGA TGC TAC TTC GGC																		
cys arg tyr val ile asp thr gly pro trp	lys ser trp pro ala gly cys tyr phe gly																		
181/61	211/71																		
CGG CGG ACT ATC AGC GGT CGC TGA GCT TCT	ACC GTG ACC AGA TC																		
arg arg thr ile ser gly arg OPA ala ser	thr val thr arg																		

FIGURE 11B'

```

1/1                               31/11
TCG CCG TCA ACG CGC TGG ACA GCA TCG GTC CCC GGT GGG TCA ATG CAC TCA TGC AGC GGT
ser pro ser thr arg trp thr ala ser val pro ala gly ser met his ser cys ser ala
61/21                               91/31
GCA ACG AAC AGC TCA ACC CTT GAA CCG GGT CCC GGC CTG CCG ACC CTC GGG CGC CGG CGT
ala thr asn ser ser thr leu glu pro gly pro gly leu pro thr leu gly arg arg arg
121/41                               151/51
GCC GCT ACG TGA TAG ACA CAG GGC CAT GGA AAT CCT GGC CAG CCG GAT GCT ACT TCG GCC
ala ala thr opa amb thr gln gly his gly asn pro gly gln pro asp ala thr ser ala
181/61                               211/71
GGC GGA CTA TCA GCG GTC GCT GAG CTT CTA CCG TGA CCA GAT C
gly gly leu ser ala val ala glu leu leu pro opa pro asp

```

FIGURE 11C'

sequence R0540c prédite par par Cole et al. Nature 393:537-541 et contenant Seg11A

[illegible]

1996, 1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 26

35/185

ORF prédite par par Cole et al. (Nature 393:537-544) et contenant Rv0546c

```

1/1                               31/11
tag tca ggg cgt gca ttc gac gac gct gta cta ccc gct ggt ggc aac tcc gat gat tgc
AMB ser gly arg ala phe asp asp ala val leu pro ala gly gly asn ser asp asp cys
61/21                               91/31
gcc gac gaa ggc cta cga cgg gct gcc ggc gct gac cgc cgc gga agc cgc cga gtg gat
ala asp glu gly leu arg arg ala ala gly ala asp arg arg gly ser arg arg val asp
121/41                               151/51
ggt cac cgc cgc cgc cac ccg acc ggt gcg gat cgc gcc tgc ggt tgc cgt cgc cgt caa
gly his arg arg pro his pro thr gly ala asp arg ala ser gly cys arg arg arg gln
181/61                               211/71
cgc gct gga cag cat cgg tcc ccg ctg ggt caa tgc act cat gca gcg ccg caa cga aca
arg ala gly gln his arg ser pro leu gly gln cys thr his ala ala pro gln arg thr
241/81                               271/91
ggt caa ccc ttg aac cgg gtc ccg gcc tgc cga ccc tgc gcc gcc gcc gcc gcc cta cgt
ala gln pro leu asn arg val pro ala cys arg pro ser ala ala gly val pro leu arg
301/101                               331/111
gat aga cac agg gcc atg gaa atc ctg gcc agc cgg atg cta ctt cgg ccg gcg gac tat
asp arg his arg ala met glu ile leu ala ser arg met leu leu arg pro ala asp tyr
361/121                               391/131
cag cgg tgc ctg agc ttc tac cgt gac cag atc ggg ctg gcg att gcc cgt gaa tac ggg
gln arg ser leu ser phe tyr arg asp gln ile gly leu ala ile ala arg glu tyr gly
421/141                               451/151
gcc gcc aca gtg ttt ttc gcc ggt cag tca ctg ctg gaa ctg gcc ggt tac gcc gag ccg
ala gly thr val phe phe ala gly gln ser leu leu glu leu ala gly tyr gly glu pro
481/161                               511/171
gac cat tgc cgg gga cat ttt ccc gcc gcg ctg tgg ctg cag gtg cgc gac ctg gag gct
asp his ser arg gly pro phe pro gly ala leu trp leu gln val arg asp leu glu ala
541/181                               571/191
acc cag acc gag ctg gtc agc cga gcc ctg tgc atc gat cga gag ccc cgt cgt gaa ccg
thr gln thr glu leu val ser arg gly val ser ile ala arg glu pro arg arg glu pro
601/201                               631/211
tcc gcc ctg cag gag atg cat ctg acc ccc cga gac ggg atc acc ctg ala ttc ctg gag
trp gly leu his glu met his val thr asp pro asp gly ile thr leu ile phe val ala
661/221                               691/231
gta ccc gat gat cag ccc ctg cat acc gcc acc ccc gat tat
val pro ala gly ile pro leu arg thr asp thr arg ala cga

```

SEQ ID N° 11F

FIGURE 11F

36/185

```

1/1                               31/11
gac cga agg gat ttc gcg act aac tcg gcc tgt aag gca acg cga ggt ctt cat gcc gag
asp arg arg asp phe ala thr asn ser ala cys lys ala thr arg gly leu his ala glu
61/21                               91/31
gac gta gac agg aag aga cag gga agc tga tga cgt cgc gta ccg gac cgc cat tct gtc
asp val asp arg lys arg gln gly ser opa opa arg arg val pro asp arg his ser val
121/41                               151/51
gag tct ttc cga gtt cag caa caa tcg aca cag aag cgg gga cca gac cgg gag gac gac
glu ser phe arg val gln gln gln ser thr gln lys arg gly pro asp arg glu asp asp
181/61                               211/71
gcg gcc cgg gcc gct tcg gcc cga gtg tct gag taa gac cag agt cac ggg tcc gtg tgt
ala ala arg ala ala ser gly arg val ser glu och asp gln ser his gly ser val cys
241/81                               271/91
gac aac cgc gcg gaa ttc aat cgg atg gcg gcc ggg acc gga ttg cgc cgg tca ccg agg
asp asn arg ala glu phe asn arg met ala gly gly thr gly leu arg arg ser pro arg
301/101
aac ctc cgg agt gat c
asn leu arg ser asp

```

SEQ ID N° 12A

FIGURE 12A

```

1/1                               31/11
acc gaa ggg att tcg cga cta act cgg cct gta ggt cca cgg agt gtc ttc atg ccg agg
thr glu gly ile ser arg leu thr arg pro val arg ala arg ala val phe met pro arg
61/21                               91/31
acg tag aca gga aga gac agg gaa gct gat gac gta gta gta gta gta gta gta gta gta
thr amb thr gly arg asp arg glu ala asp arg val ala thr arg thr ala ile leu ser
121/41                               151/51
agt ctt tcc gag ttc agc aac aat cga cac aat arg ggt ggt ggt ggt ggt ggt ggt ggt
ser leu ser ala phe ser asn asn arg his arg ser ala arg ala thr gly arg thr thr
181/61                               211/71
gga cca gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa
arg pro gly pro leu arg ala ala ala leu ser arg arg arg arg arg arg arg arg arg
241/81                               271/91
aca aat gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa
thr thr ala arg asn ser ala gly thr arg ala arg arg arg arg arg arg arg arg arg
301/101
aac ttc gaa gaa gaa
asn ser ala val ala

```

SEQ ID N° 12B

```

1/11
ccg aag gga ttt cgc gac taa ctc ggc ctg taa ggc aac gcg agg tct tca tgc cga gga
pro lys gly phe arg asp OCH leu gly leu OCH gly asn ala arg ser ser cys arg gly
61/21
cgt aga cag gaa gag aca ggg aag ctg atg acg tgc cgt acc gga cgg cca ttc tgt cga
arg arg gln glu glu thr gly lys leu met thr ser arg thr gly pro pro phe cys arg
121/41
gtc ttt ccg agt tca gca aca atc gac aca gaa cgc ggg acc aga cgg gga gga cga cgc
val phe pro ser ser ala thr ile asp thr glu ala gly thr arg pro gly gly arg arg
181/61
ggc ccg ggc cgc ttc ggg ccg agt gtc tga gta aga cca gag tca cgg gtc cgt gtg tga
gly pro gly arg phe gly pro ser val OPA val arg pro glu ser arg val arg val OPA
241/81
caa ccg cgc gga att caa tgc gat ggc ggg cgg gac cgg att gcg ccg gtc acc gag gaa
gln pro arg gly ile gln ser asp gly gly arg asp arg ile ala pro val thr glu glu
301/101
ccg ccg gag tga tc
pro pro glu OPA

```

FIGURE 12C

[illegible]

38/185

32/11  
 GGA TTT CGT TGC CCG ATG GAT TGT TTG TAC GGT TTG GGA AAA ACA CTT GAA GTC CTT TTT  
 gly phe arg cys pro met asp cys leu tyr gly leu gly lys thr leu glu val leu phe  
 62/21  
 ATT GGC AAT GCT GGA AAT GGA CAT TCC AAT ATT GCG CGA ATT AAC CGA ACA CGG TGA GGG  
 ile gly asn ala gly asn gly his ser asn ile ala arg ile asn arg thr arg OPA gly  
 122/41  
 GGG GGC AAG CGT TTG TAC CCG GGC CAG CAA GCG CCG CCG ACC GGT TGA CCG AAG CCA GCA  
 gly gly lys arg leu tyr arg gly gln gln ala pro pro thr gly OPA pro lys pro ala  
 182/61  
 TGT TGT TGT GTC AGC GCG GGC TTG GTC TCG ATG TCG CCG CCT TGG CTG GAC CCG CTT CTT  
 cys cys cys val ser ala gly leu val ser met ser arg pro trp leu asp pro leu leu  
 242/81  
 CAA AAC AGG TTG AAC TTA ACG ACT CAA GAA GCG AAA CCG TTG AAC CCG CAC GTC GCT CCG  
 gln asn arg leu asn leu thr thr gln glu arg lys arg leu asn arg asp val ala pro  
 302/101  
 GAC ACC AAT TTG ACT CCG CTC TTT GGC AAT TGA AGG TGA GCT GCG AGC AGC CCG GTG ACC  
 asp thr asn leu thr arg leu phe gly asn OPA arg OPA ala ala ser ser arg val thr  
 362/121  
 GCA TCG TTG GCG TTG CCA TCA ATC GCG GCG TCG CCG ACG TAG ATA ATC AGC TCA CCG TTG  
 ala ser leu ala leu pro ser ile ala gly ser arg thr AMB ile ile ser ser pro leu  
 422/141  
 GGA CCG ACC TCG ACC AGG GGT CCT TTG TGA CTG CCG GCG TTG ACG CCG ACG ACC ACA GAG  
 gly pro thr ser thr arg gly pro leu OPA leu pro gly leu thr arg thr thr thr glu  
 482/161  
 TCG GTC ATC GCG TAA GGC TAC CGT TCT GAC CTG GGG CTG CGT GGG CCG CGA CGA CGT GAG  
 ser val ile ala OCH gly tyr arg ser asp leu gly leu arg gly arg arg arg arg glu  
 542/181  
 GCA CGT CAT GTC TCA GCG GCC CAC CCG CAC CTC GGT CCG CCG CAG TAT GTC ACC ATG TGC  
 ala arg his val ser ala ala his arg his leu gly arg arg gln tyr val ser met cys  
 602/201  
 AGA TGA CTC CAC GCA GCG TTG TTC GCA TCG TTG GTG TCG TCG TTG CGA CGA CCT TCG CCG  
 arg OPA leu his ala ala leu phe ala ser leu val ser trp leu arg arg pro trp arg  
 662/221  
 TGT TGA GCG CAC CCG CCG GCG GTC GTG CCG CCG ATG CCG ATT  
 trp OPA ala his pro pro ala val val pro arg met arg ile

SEQ ID N° 13B

FIGURE 13B

39/185

33/11  
 GAT TTC GTT GCC CGA TGG ATT GTT TGT ACG GTT TGG GAA AAA CAC TTG AAG TCC TTT TTA  
 asp phe val ala arg trp ile val cys thr val trp glu lys his leu lys ser phe leu  
 63/21  
 TTG GCA ATG CTG GAA ATG GAC ATT CCA ATA TTG CGC GAA TTA ACC GAA CAC GGT GAG GGG  
 leu ala met leu glu met asp ile pro ile leu arg glu leu thr glu his gly glu gly  
 123/41  
 GGG GCA AGC GTT TGT ACC GGG GCC AGC AAG CGC CGC CGA CCG GTT GAC CGA AGC CAG CAT  
 gly ala ser val cys thr gly ala ser lys arg arg arg pro val asp arg ser gln his  
 183/61  
 GTT GTT GTG TCA GCG CGG GCT TGG TCT CGA TGT CCC GGC CTT GGC TGG ACC CGC TTC TTC  
 val val val ser ala arg ala trp ser arg cys pro gly leu gly trp thr arg phe phe  
 243/81  
 AAA ACA GGT TGA ACT TAA CGA CTC AAG AAC GGA AAC GGT TGA ACC GCG ACG TCG CTC CGG  
 lys thr gly opa thr och arg leu lys asn gly asn ala opa thr ala thr ser leu arg  
 303/101  
 ACA CCA ATT TGA CTC GGG TCT TTG GCA ATT GAA GST GAG CTG CGA GCA GCC GGG TGA CCG  
 thr pro ile opa leu gly ser leu ala ile glu gly glu leu arg ala ala gly opa pro  
 363/121  
 CAT CGT TGG CCT TGC CAT CAA TCG CCG GCT CGC GGA CGT AGA TAA TCA GGT CAC CGT TGG  
 his arg trp pro cys his gln ser pro ala arg gly arg arg och ser ala his arg trp  
 423/141  
 GAC CGA CCT CGA CCA GGG GTC CTT TST GAC TGC CCG SCT TGA CGC GGA CGA CCA CAG AGT  
 asp arg pro arg pro gly val leu cys asp cys arg ala opa arg gly arg pro gln ser  
 483/161  
 CGG TCA TCG CCT AAG SCT ACC GTT CTG ACC TGG GGC TGC CTG GGC GCC GAC GAC GTG AGG  
 arg ser ser pro lys ala thr val leu thr trp gly cys val gly ala asp asp val arg  
 543/181  
 CAC GTC ATG TCT CAG CGG CGC ACC SCC ACC TGG GTC GCG SCC AGT ATG TCA GCA TGT GCA  
 his val met ser gln arg pro thr ala thr ser val ala gly ser met ser ala cys ala  
 603/201  
 GAT GAC TCC ACG CAG CCT TGT TCG CTT CCT TCG TGT CGT GGT TCG GAC GAC CTT GGT GCT  
 asp asp ser thr gln pro cys ser his arg trp cys arg gly cys asp asp leu gly ala  
 663/221  
 GGT GAG CGC ACC CGC CGG CGG TGG TGC CGC GCA TGC GGA TC  
 gly glu arg thr arg arg arg ser cys arg ala cys gly

SEQ ID N° 13C

FIGURE 13C

40/185

partie de la séquence nucléotidique de seq13A

1/1 31/11  
 GGG TCC TTT GTG ACT GCC GGG CTT GAC GCG GAC CAC AGA GTC GGT CAT CGC CTA AGG  
 gly ser phe val thr ala gly leu asp ala asp asp his arg val gly his arg leu arg  
 61/21 91/31  
 CTA CCG TTC TGA CCT GGG GCT GCG TGG GCG CCG ACG ACG TGA GGC ACG TCA TGT CTC AGC  
 leu pro phe OPA pro gly ala ala trp ala pro thr thr OPA gly thr ser cys leu ser  
 121/41 151/51  
 GGC CCA CCG CCA CCT CGG TCG CCG GCA GTA TGT CAG CAT GTG CAG ATG ACT CCA CGC AGC  
 gly pro pro pro pro arg ser pro ala val cys gln his val gln met thr pro arg ser  
 181/61 211/71  
 CTT GTT CGC ATC GTT GGT GTC GTG GTT GCG ACG ACC TTG GCG CTG GTG AGC GCA CCC GGC  
 leu val arg ile val gly val val val ala thr thr leu ala leu val ser ala pro ala  
 241/81  
 GGC GGT CGT GCC GCG CAT GCG GAT C  
 gly gly arg ala ala his ala asp

SEQ ID N° 13A'

FIGURE 13A'

1/1 31/11  
 GGT CCT TTG TGA CTG CCG GGC TTG ACG CGG ACG ACC ACA GAG TCG CTC ATC GCC TAA GGC  
 gly pro leu OPA leu pro gly leu thr arg thr thr thr glu ser val ile ala OCH gly  
 61/21 91/31  
 TAC CGT TCT GAC CTG GGG CTG CGT GGG GCG CGA CGA CGT GAG GCA CGT CAT GTC TCA CCG  
 tyr arg ser asp leu gly leu arg gly arg arg arg arg glu ala arg his val ser ala  
 121/41 151/51  
 GCC CAC CGC CAC CTC CGT CGC CGC CAG TAT CTC ACC ATG TGC AGA TGA CTC CAC GCA GGC  
 ala his arg his leu gly arg arg gln tyr val ser met cys arg OPA leu his ala ala  
 181/61 211/71  
 TTG TTC GGA TCG TTG GTG TCG TGG TTG CGA CGA CCT TGG GCG TGG TGA GCG CAC CGC CGG  
 leu phe ala ser leu val ser trp leu arg arg pro trp arg trp OPA ala his pro pro  
 241/81  
 GCG GTC GTG CCG CGC ATG CGG ATC  
 ala val val pro arg met arg ile

SEQ ID N° 13B'

FIGURE 13B'

1/1 31/11  
 GGT CCT TTG TGA CTG CCG GGC TTG ACG CGG ACG ACC ACA GAG TCG CTC ATC GCC TAA GGC  
 val leu tyr asp cys arg ala ala arg gly arg arg arg arg arg arg arg arg arg arg  
 61/21 91/31  
 AAC GTT GAT AAC TGG GAT TAT GTC GCG GCG GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 thr val leu thr thr trp gly cys val gly ala arg arg arg arg arg arg arg arg arg  
 121/41 151/51  
 TTT AAT GAT AAT TTT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ptt thr ala thr ser val ala gly ser ptt ptt ptt ptt ptt ptt ptt ptt ptt ptt ptt

SEQ ID N° 13C'

REPLACEMENT SHEET (RULE 26)



41/185

séquence Rv1984c prédite par Cole et al. (Nature 393:537-544) et contenant seq13A'

```

1/1                               31/11
atg act cca cgc agc ctt gtt cgc atc gtt ggt gtc gtg gtt gcg acg acc ttg gcg ctg
Met thr pro arg ser leu val arg ile val gly val val val ala thr thr leu ala leu
61/21                               91/31
gtg agc gca ccc gcc gcc ggt cgt gcc gcg cat gcg gat ccg tgt tgc gac atc gcg gtc
val ser ala pro ala glv gly arg ala ala his ala asp pro cys ser asp ile ala val
121/41                               151/51
gtt ttc gct cgc gcc acg cat cag gct tct ggt ctt gcc gac gtc ggt gag gcg ttc gtc
val phe ala arg gly thr his gln ala ser gly leu gly asp val gly glu ala phe val
181/61                               211/71
gac tgc ctt acc tgc caa gtt gcc gcc cgg tgc att ggg gtc tac gcg gtg aac tac cca
asp ser leu thr ser gln val gly gly arg ser ile gly val tyr ala val asn tyr pro
241/81                               271/91
gca agc gac gac tac cgc gcg agc gcg tca aac ggt ttc gat gat gcg agc gcc cac atc
ala ser asp asp tyr arg ala ser ala ser asn gly ser asp asp ala ser ala his ile
301/101                               331/111
cag cgc acc gtc gcc agc tgc ccg aac aac agg att gtg ctt ggt gcc tat tgc cag ggt
gln arg thr val ala ser cys pro asn thr arg ile val leu gly gly tyr ser gln gly
361/121                               391/131
gcg acg gtc atc gat ttg tcc acc tgc gcg atg ccg ccc gcg gtg gca gat cat gtc gcc
ala thr val ile asp leu ser thr ser ala met pro pro ala val ala asp his val ala
421/141                               451/151
gct gtc gcc ctt ttc gcc gag cca tcc agt ggt ttc tcc agc atg ttg tgg gcc gcc gcc
ala val ala leu phe gly glu pro ser ser gly phe ser ser met leu trp gly gly gly
481/161                               511/171
tgc ttg ccg aca atc ggt ccg ctg tat agc tct aac acc ala aac ttg tgt gct acc gac
ser leu pro thr ile gly pro leu tyr ser ser lys thr ile asn leu cys ala pro asp
541/181                               571/191
gat cca ata tgc acc gca gcc gcc aat att atg gcc cat gtt tgc tat gtt cag tca gcc
asp pro ile cys thr gly gly gly asn ile met ala his val ser tyr val gln ser gly
601/201                               631/211
atg aca agc cag gcg gcg aca tcc gcg gcg aac agg ctg gat cac gcc gcc tga
met thr ser gln ala ala thr phe ala ala aac arg leu asp his ala gly opa

```

SEQ ID N° 133

FIGURE 13D

42/185

Seq13F: ORF prédite par Cole et al. (Nature 393:537-544) et contenant Rv1984c

```

1/1                               31/11
tga ggc acg tca tgt ctc agc ggc cca ccg cca cct cgg tcg ccg gca gta tgt cag cat
OPA gly thr ser cys leu ser gly pro pro pro pro arg ser pro ala val cys gln his
61/21                               91/31
gtg cag atg act cca cgc agc ctt gtt cgc atc gtt ggt gtc gtg gtr gcg acg acc ttg
val gln met thr pro arg ser leu val arg ile val gly val val val ala thr thr leu
121/41                               151/51
gcg ctg gtg agc gca ccc gcc ggc ggt cgt gcc gcg cat gcg gat ccg tgt tcg gac atc
ala leu val ser ala pro ala gly gly arg ala ala his ala asp pro cys ser asp ile
181/61                               211/71
gcg gtc gtt ttc gct cgc ggc agc cat cag gct tct ggt ctt ggc gac gtc ggt gag gcg
ala val val phe ala arg gly thr his gln ala ser gly leu gly asp val gly glu ala
241/81                               271/91
ttc gtc gac tcg ctt acc tcg caa gtt ggc ggg cgg tcg att ggg gtc tac gcg gtg aac
phe val asp ser leu thr ser gln val gly gly arg ser ile gly val tyr ala val asn
301/101                               331/111
tac cca gca agc gac gac tac cgc gcg agc gcg tca aac ggt tcc gat gat gcg agc gcc
tyr pro ala ser asp asp tyr arg ala ser ala ser asn gly ser asp asp ala ser ala
361/121                               391/131
cac atc cag cgc acc gtc gcc agc tgc ccg aac acc agg att gtg ctt ggt gcc tat tcg
his ile gln arg thr val ala ser cys pro asn thr arg ile val leu gly gly tyr ser
421/141                               451/151
cag ggt gcg acg gtc atc gat ttg tcc acc tcg gcg atg ccg ccc gcg gtg gca gat cat
gln gly ala thr val ile asp leu ser thr ser ala met pro pro ala val ala asp his
481/161                               511/171
gtc gcc gct gtc gcc ctt ttc ggc gag cca tcc aqt ggt ttc tcc agc atg ttg tgg gcc
val ala ala val ala leu phe gly glu pro ser ser gly phe ser ser met leu trp gly
541/181                               571/191
ggc ggg tcg ttg ccg aca atc ggt ccg ctg tat agc tat aag acc ata aac ttg tgt gct
gly gly ser leu pro thr ile gly pro leu tyr ser ser lys thr ile asn leu pys ala
601/201                               631/211
ccc gac gat cca ata tgc acc gga ggc ggc aat att atg gcg cat gtt tcg tat gtt cag
pro asp asp pro ile cys thr gly gly gly asn ile met ala his val ser tyr val ala
661/221                               691/231
tgg ggg atg aca agc cag ggt ggt acc ttc ser cag aac agc ttc att cag ggt ggt ttc
ser gly met thr ser ala ala ala thr phe ala ala asn ala leu asp his ala gly cpa

```

SEQ ID N° 13F

FIGURE 13F

43/185

31/11  
 CCA CCG GGG CTG GAG GGG CGA ATG TGC GCC GAA CGC CGT CGG CCA ACT TGG CCG CTG AGG  
 pro pro gly leu glu gly arg met cys ala glu arg arg arg pro thr trp pro leu arg  
 61/21  
 GCG GCT GAT CCC CTG GCC CGA GAC GGG GCA AGC CAA TAG CGG CTC CAT CGG GCT TTG CTG  
 ala ala asp pro leu ala arg asp gly ala ser gln AMB arg leu his arg ala leu leu  
 121/41  
 GTA GCG GTT CGG CGG GAA CCG AGC GCC GAC GTT GTC GGT GCC CGG TGA TAT ATT GGG TCA  
 val ala val arg arg glu pro ser ala asp val val gly ala arg OPA tyr ile gly ser  
 181/61  
 GAC GGG TAT GGC GGC GAC TGA GGT GAT CTG CGA CAC GCC GCC GCG GTG CTC GAG CCA GGC  
 asp gly tyr gly gly asp OPA gly asp leu arg his ala ala ala val leu glu pro gly  
 241/81  
 TTA CGA CCA GGG AAT TTC GAA AAT GTT ATT CAC AAC ATC TTG TAT CTC TTC CTC GGT GCC  
 leu arg pro gly asn phe glu asn val ile gln asn ile leu tyr leu phe leu arg ala  
 301/101  
 ACC CCC TAG GTG TAG TGT TTT CGA GTA CCG GCA GAT CCC AGT TCA CCA GTC TCA CCA GAT  
 thr pro AMB val AMB cys phe arg val pro ala asp pro ser ser pro val ser pro asp

C

SEQ ID N° 14A

FIGURE 14A

32/11  
 CAC CGG GGC TGG AGG GGC GAA TGT CGC CGG AAC GGC GTT GGC CAA GTT GGC CGC TCA GGG  
 his arg gly trp arg gly glu cys ala pro asn ala val gly gln leu gly arg OPA gly  
 62/21  
 CGG CTG ATC CCG TGG CCG GAG AGC GGG CAA GCG AAT AGC GGC TCG ATC GGG GTT TGT TGG  
 arg leu ile pro trp pro glu thr gly ala ala asn ser gly ser ile gly leu cys trp  
 122/41  
 TAG CGG TTC GGC GGC AAC CGA GCG CCG AGC TTG TCG GTG CCG GGT GAT ATA TTG GGT CAG  
 AMB arg phe gly gly asn arg ala pro thr leu ser val pro gly trp ile leu gly ala  
 182/61  
 ACG GGT ATG GCG GCG AGT GAG GCG ATC TCG GAG AGC CCG GCG TCG TCG AGC CAG GGT  
 thr gly met ala ala thr glu val ile cys asp thr pro pro arg tyr ser ser ala ala  
 242/81  
 TAG GAG CAG GGA ATT TCG AAA ATC TTA TTC AGA AAT TCG TGT ATT TGT TGT TGT TGT TGT  
 tyr asp gln gly ile ser lys met leu phe arg thr ser cys ile ser ser ser val pro  
 302/101  
 GGC GGT AGC TGT AGT GTT TTC GAG TAC CAG CAG ATC TCA GTT CAG CAG TGT CAG TAC ATC  
 pro pro arg thr ser val phe gln tyr arg ala ile pro val his ala ser ala thr ile

44/185

33/11  
 ACC GGG GCT GGA GGG GCG AAT GTG CGC CGA ACG CCG TCG GCC AAC TTG GCC GCT GAG GGC  
 thr gly ala gly gly ala asn val arg arg thr pro ser ala asn leu ala ala glu gly  
 63/21  
 GGC TGA TCC CCT GGC CCG AGA CGG GGC AAG CCA ATA GCG GCT CCA TCG GGC TTT GCT GGT  
 gly OPA ser pro gly pro arg arg gly lys pro ile ala ala pro ser gly phe ala gly  
 123/41  
 AGC GGT TCG GCG GGA ACC GAG CGC CGA CGT TGT CGG TGC CCG GTG ATA TAT TGG GTC AGA  
 ser gly ser ala gly thr glu arg arg arg cys arg cys pro val ile tyr trp val arg  
 183/61  
 CGG GTA TGG CGG CGA CTG AGG TGA TCT GCG ACA CGC CGC CCG GGT GCT CGA GCC AGG CTT  
 arg val trp arg arg leu arg OPA ser ala thr arg arg arg gly ala arg ala arg leu  
 243/81  
 ACG ACC AGG GAA TTT CGA AAA TGT TAT TCA GAA CAT CTT GTA TCT CTT CCT CCG TGC CAC  
 thr thr arg glu phe arg lys cys tyr ser glu his leu val ser leu pro pro cys his  
 303/101  
 CCC CTA GGT GTA GTG TTT TCG AGT ACC GGC AGA TCC CAG TTC ACC AGT CTC ACC AGA TC  
 pro leu gly val val phe ser ser thr gly arg ser gln phe thr ser leu thr arg

SEQ ID N° 14C

FIGURE 14C

partie de la séquence nucléotidique de seq14A

1/1 31/11  
 TTT TCG AGT ACC GGC AGA TCC CAG GTT CAG CAG GTC TCA CCA GAT C  
 phe ser ser thr gly arg ser gln val his gln val ser pro asp

SEQ ID N° 14A'

FIGURE 14A'

1/1 31/11  
 TGT TTT CGA GTA GGT GGA GAT CCG AGG GTT AAT AAT TTT TTT TTT AAT  
 ys phe arg val pro ala asp pro arg phe thr thr ser thr thr thr

SEQ ID N° 14C

FIGURE 14C

SEQ ID N° 14C

FIGURE 14C  
 SEQ ID N° 14C

45/185

ORF prédite d'après la séquence publiée par Cole et al. (Nature 393:537-544) et contenant seq14A'

```

1/1                                31/11
TAG CGG TTC GGC GGG AAG CTA GCG GCG ACG TTG TCG GTG GCC GGT GAT ATA TTG GGT CAG
AMB arg phe gly gly lys leu ala ala thr leu ser val ala gly asp ile leu gly gln
61/21                                91/31
ACG GGT ATG GCG GCG GCT GAG GTG ATC TGC GAC ACG CCG CCG CGG TGC TCG AGC CAG GCT
thr gly met ala ala ala glu val ile cys asp thr pro pro arg cys ser ser gln ala
121/41                                151/51
TAC GAC CAG GGA ATT TCG AAA ATG TTA TTC AGA ACA TCT TGT ATC TCT TCT CCG TGC CAC
tyr asp gln gly ile ser lys met leu phe arg thr ser cys ile ser ser pro cys his
181/61                                211/71
CCC CTA GGT GTA GTG TTT TCG AGT ACC GGC AGA TCC CAG GTT CAC CAG GTC TCA CCA gat
pro leu gly val val phe ser ser thr gly arg ser gln val his gln val ser pro asp
241/81                                271/91
cca cgg ggc gcg atg aac ttc ccg gca tcg gca tcg cca ggt cga cgg acg tgg tcg cgc
pro arg gly ala met asn phe pro ala ser ala ser pro gly arg arg thr trp ser arg
301/101                                331/111
tat gac ggg aat ctg gag cct tgt cgg gcc gct cca cat atc gaa gat gca cta ctt gag
tyr asp gly asn leu glu pro cys arg ala ala gln his ile glu asp ala leu leu glu
361/121                                391/131
tcg ttg cca gat cct gtc aga ttc ccg att tcc gca aag gag cgg tac gcc cat gac cgt
ser leu pro asp pro val arg phe pro ile ser ala lys glu arg-tyr ala his asp arg
421/141
gac cgt tta cac taa
asp arg leu his OCH

```

SEQ ID N° 14F

FIGURE 14F

46/185

Séquence RV3054c prédite par Cole et l. (Nature 393:537-544)  
pouvant être en phase avec Seq14A'

```

1/1                               31/11
gtg tca gat acc aag tcc gac atc aaa atc ttg gcc tta gtg gga agc ctg cgc gcg gcg
val ser asp thr lys ser asp ile lys ile leu ala leu val gly ser leu arg ala ala
61/21                               91/31
tcg ttc aac cgc cag atc gcc gag ctg gct gcc aag gtc gct ccg gac ggc gtc acc gtc
ser phe asn arg gln ile ala glu leu ala ala lys val ala pro asp gly val thr val
121/41                               151/51
acc atg ttc gag ggg ctg ggg gac ctg ccg ttc tac aac gaa gac atc gac aca gcg acg
thr met phe glu gly leu gly asp leu pro phe tyr asn glu asp ile asp thr ala thr
181/61                               211/71
gag gtg ccg gcg ccg gtg agc gcg ttg ccg gag gcc gcg tct gac gcg cac gct gcc ttg
glu val pro ala pro val ser ala leu arg glu ala ala ser asp ala his ala ala leu
241/81                               271/91
gtg gtc acg ccg gaa tac aac ggc agc att ccg gcc gtg atc aag aac gcg atc gac tgg
val val thr pro glu tyr asn gly ser ile pro ala val ile lys asn ala ile asp trp
301/101                               331/111
ctg tcc agg cca ttc ggc gat ggc gcg ttg aag gac aag ccg ttg gcc gtg atc ggc ggc
leu ser arg pro phe gly asp gly ala leu lys asp lys pro leu ala val ile gly gly
361/121                               391/131
tcc atg ggc cgc tac ggc ggg gta tgg gcg cac gac gag act cgc aag tcg ttc agc atc
ser met gly arg tyr gly gly val trp ala his asp glu thr arg lys ser phe ser ile
421/141                               451/151
gct ggc acg cgg gtg gtc gat gcg atc aaa ctg tcg gtg ccg ttc caa act ctg ggc aag
ala gly thr arg val val asp ala ile lys leu ser val pro phe gln thr leu gly lys
481/161                               511/171
tcg gtc gcg gac gac gcc ggg ctg gcg gcg aat gtg cgc gac gcc gtc ggc aac ttg gcc
ser val ala asp asp ala gly leu ala ala asn val arg asp ala val gly asn leu ala
541/181
gct gag gtc ggc tga
ala glu val gly opa

```

SEQ ID N° 14R

FIGURE 14R

47/185

ORF prédite par Cole et al. (Nature 393:537-544) et contenant Rv3054c

```

1/1                                31/11
taa cgc gat cgg aat aaa tcg gac cat ggt ccg gtt ggc tcg tgc aag gac gtg gac caa
OCH arg asp arg asn lys ser asp his gly pro val gly ser cys lys asp val asp gln
61/21                                91/31
caa gcg gaa agg aac gta gca gtg tca gat acc aag tcc gac atc aaa atc ttg gcc tta
gln ala glu arg asn val ala val ser asp thr lys ser asp ile lys ile leu ala leu
121/41                                151/51
gtg gga agc ctg cgc gcg gcg tcg ttc aac cgc cag atc gcc gag ctg gct gcc aag gtc
val gly ser leu arg ala ala ser phe asn arg gln ile ala glu leu ala ala lys val
181/61                                211/71
gct ccg gac gcc gtc acc gtc acc atg ttc gag ggg ctg ggg gac ctg ccg ttc tac aac
ala pro asp gly val thr val thr met phe glu gly leu gly asp leu pro phe tyr asn
241/81                                271/91
gaa gac atc gac aca gcg acg qag gtg ccg gcg ccg gtg agc gcg ttg cgg gag gcc gcg
glu asp ile asp thr ala thr glu val pro ala pro val ser ala leu arg glu ala ala
301/101                               331/111
tgt gac gcg cac gct gcc ttg gtg gtc acg ccg gaa tac aac gcc agc att ccg gcc gtg
ser asp ala his ala ala leu val val thr pro glu tyr asn gly ser ile pro ala val
361/121                               391/131
atc aag aac gcg atc gac tgg ctg tcc agg cca ttc gcc gat gcc gcg ttg aag gac aag
ile lys asn ala ile asp trp leu ser arg pro phe gly asp gly ala leu lys asp lys
421/141                               451/151
ccg ttg gcc gtg atc gcc gcc tcc atg gcc ccg tac gcc ggg gta tgg gcg cac qac gag
pro leu ala val ile gly gly ser met gly arg tyr gly gly val trp ala his asp glu
481/161                               511/171
act cgc aag tcg ttc agc atc gct gcc acg ccg gtg gtc gat gcg atc aaa ctg tcg gtg
thr arg lys ser phe ser ile ala gly thr arg val val asp ala ile lys leu ser val
541/181                               571/191
ccg ttc caa act ctg gcc aag tcg gtc gcg gac gac gcc ggg ctg gcg gcg aat gtg cgc
pro phe gln thr leu gly lys ser val ala asp asp ala gly leu ala ala asn val arg
601/201                               631/211
gac gcc gtc gcc aac ttg gct gct gag gtc gcc tga
asp ala val gly asn leu ala ala glu val gly OPA

```

SEQ ID N° 14P

FIGURE 14P

48/185

fragment d'après la séquence publiée par Cole et al. (Nature 393:537-544) et contenant seq 14F' et seq 14P'

1/1 31/11  
 taa cgc gat cgg aat aaa tcg gac cat ggt cgg gtt ggc tcg tgc aag gac gtg gac caa  
 OCH arg asp arg asn lys ser asp his gly pro val gly ser cys lys asp val asp gln  
 asn ala ile gly ile asn arg thr met val arg leu ala arg ala arg thr trp thr asn  
 thr arg ser glu OCH ile gly pro trp ser gly trp leu val gln gly arg gly pro thr  
 61/21 91/31  
 caa cgc gaa agg aac gta gca gtg tca gat acc aag tcc gac atc aaa atc ttg gcc tta  
 gln ala glu arg asn val ala val ser asp thr lys ser asp ile lys ile leu ala leu  
 lys arg lys gly thr AMB gln cys gln ile pro ser pro thr ser lys ser trp pro AMB  
 ser gly lys glu arg ser ser val arg tyr gln val arg his gln asn leu gly leu ser  
 121/41 151/51  
 gtg gga agc ctg cgc ggc ggc tcg ttc aac cgc cag atc gcc gag ctg gct gcc aag gtc  
 val gly ser leu arg ala ala ser phe asn arg gln ile ala glu leu ala ala lys val  
 trp glu ala cys ala arg arg arg ser thr ala arg ser pro ser trp leu pro arg ser  
 gly lys pro ala arg gly val val gln pro pro asp arg arg ala gly cys gln gly arg  
 181/61 211/71  
 gct cgc gac gcc gtc acc gtc acc atg ttc gag ggg ctg ggg gac ctg cgc ttc tac aac  
 ala pro asp gly val thr val thr met phe glu gly leu gly asp leu pro phe tyr asn  
 leu arg thr ala ser pro ser pro cys ser arg gly trp gly thr cys arg ser thr thr  
 ser gly arg arg his arg his his val arg gly ala gly gly pro ala val leu gln arg  
 241/81 271/91  
 gaa gac atc gac aca gcg acg gag gtg cgc ggc cgc gtg agc gcg ttg cgg gag gcc gcg  
 glu asp ile asp thr ala thr glu val pro ala pro val ser ala leu arg glu ala ala  
 lys thr ser thr gln arg arg arg cys arg arg arg OPA ala arg cys gly arg pro arg  
 arg his arg his ser asp gly gly ala gly ala gly glu arg val ala gly gly arg val  
 301/101 331/111  
 tct gac gcg cac gct gcc ttg gtg gtc acg cgc gaa tac aac gcc agc att cgc gcc gtg  
 ser asp ala his ala ala leu val val thr pro glu tyr asn gly ser ile pro ala val  
 leu thr arg thr leu pro trp trp ser arg arg asn thr thr ala ala phe arg pro OPA  
 OPA arg ala arg cys leu gly gly his ala gly ile gln arg gln his ser gly arg asp  
 361/121 391/131  
 atc aag aac gcg atc gac tgg ctg tcc agg cca ttc gcc gat ggc gag ttg aag gac aag  
 ile lys asn ala ile asp trp leu ser arg pro phe gly asp gly ala leu lys asp lys  
 ser arg thr arg ser thr gly cys pro gly his ser ala met ala arg OPA arg thr ser  
 gln glu arg asp arg leu ala val gln ala ile arg arg trp arg val glu gly gln ala  
 421/141 451/151  
 ccg ttg gcc gtg atc gcc gcc tcc atg gcc gcc tac gtc gcc gta tgg ggc gac gac gac  
 pro leu ala val ile gly gly ser met gly arg tyr gly gly val trp ala his asp glu  
 arg trp pro OPA ser ala ala pro trp ala ala thr ala gly tyr gly arg thr thr arg  
 val gly arg asp arg arg leu his gly pro leu arg arg gly met gly ala arg arg asp  
 481/161 511/171  
 act cgc aag tcc ttc aac atc gct gcc acc gcc ggc gtc gta gat gag atc aaa ctg tgg gtc  
 thr arg lys ser phe ser ile ala gly thr arg val val asp ala ile lys leu ser val  
 leu ala ser asp ser ala ser leu ala arg gly trp ser met arg ser asn ggc arg ggc  
 ser gln val val gln his arg trp his ala gly gly arg cys arg gln thr val gly ala  
 541/181 571/191  
 arg ttc aac act ctg gcc aag tgg gtc gcc gac gac gac gac gac gac gac gac gac gac  
 pro phe gln thr leu gly lys ser val ala asp asp ala gly leu ala ala asn val arg  
 arg ser lys leu trp ala ser arg ser arg thr thr thr gly trp arg arg met ggc ala  
 val pro asn ser gly gln val gly arg gly arg arg arg ala arg gly gly cys ala arg  
 601/201 631/211  
 aa gtc gtc aa gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc  
 asp ala val gly asp leu ala ala ala val gly aa ser leu gly arg gly gly ser ala  
 thr pro ser ala thr trp trp thr arg ser ala arg arg arg arg arg arg arg arg arg

SEQ ID NO: 110

REEMPLACEMENT REGIE 26



49/185

721/241 751/251  
 cgg tgg ccg gtg ata tat tgg gtc aga cgg gta tgg cgg cgg ctg agg tga tct gcg aca  
 arg trp pro val ile tyr trp val arg arg val trp arg arg leu arg OPA ser ala thr  
 gly gly arg OPA tyr ile gly ser asp gly tyr gly gly gly OPA gly asp leu arg his  
 val ala gly asp ile leu gly gln thr gly met ala ala ala glu val ile cys asp thr  
 781/261 811/271  
 cgc cgc cgc ggt gct cga gcc agg ctt acg acc agg gaa ttt cga aaa tgt tat tca gaa  
 arg arg arg gly ala arg ala arg leu thr thr arg glu phe arg lys cys tyr ser glu  
 ala ala ala val leu glu pro gly leu arg pro gly asn phe glu asn val ile gln asn  
 pro pro arg cys ser ser gln ala tyr asp gln gly ile ser lys met leu phe arg thr  
 841/281 871/291  
 cat ctt gta tct ctt ctc cgt gcc acc ccc tag gtg tag tgt ttt cga gta ccg gca gat  
 his leu val ser leu leu arg ala thr pro AMB val AMB cys phe arg val pro ala asp  
 ile leu tyr leu phe ser val pro pro pro arg cys ser val phe glu tyr arg gln ile  
 ser cys ile ser ser pro cys his pro leu gly val val phe ser ser thr gly arg ser  
 901/301 931/311  
 ccc agg ttc acc agg tct cac cag atc cac ggg gcg cga tga act tcc cgg cat cgg cat  
 pro arg phe thr arg ser his gln ile his gly ala arg OPA thr ser arg his arg his  
 pro gly ser pro gly leu thr arg ser thr gly arg asp glu leu pro gly ile gly ile  
 gln val his gln val ser pro asp pro arg gly ala met asn phe pro ala ser ala ser  
 961/321 991/331  
 cgc cag gtc gac gga cgt ggt cgc gct atg acg gga atc tgg agc ctt gtc ggg ccg ctc  
 arg gln val asp gly arg gly arg ala met thr gly ile trp ser leu val gly pro leu  
 ala arg ser thr asp val val ala leu OPA arg glu ser gly ala leu ser gly arg ser  
 pro gly arg arg thr trp ser arg tyr asp gly asn leu glu pro cys arg ala ala gln  
 1021/341 1051/351  
 aac ata tcg aag atg cac tac ttg agt cgt tgc cag atc ctg tca gat tcc cga ttt ccg  
 asn ile ser lys met his tyr leu ser arg cys gln ile leu ser asp ser arg phe pro  
 thr tyr arg arg cys thr thr OPA val val ala arg ser cys gln ile pro asp phe arg  
 his ile glu asp ala leu leu glu ser leu pro asp pro val arg phe pro ile ser ala  
 1081/361 1111/371  
 caa agg agc ggt acg ccc atg acc gtg acc gtt tac act aa  
 gln arg ser gly thr pro met thr val thr val tyr thr  
 lys gly ala val arg pro OPA pro OPA pro phe thr leu  
 lys glu arg tyr ala his asp arg asp arg leu his OCH

SEQ ID N° 14Q(suite)

FIGURE 14Q(suite)

1/1 31 11  
 CAA GCG CGG GCG CGA GCG TTT GCG GTT TTT GCG GCG TTA GCG GAA GAA GAA GCG GCG GCG  
 gln ala arg pro arg leu phe ala val leu gly leu leu pro ala his his leu ala arg  
 01/21 31/31  
 GCG GAC GAT GGT GTG GAT GAG TTT GCG TTT GTT GGT GCG GCG GCG GCG GCG GCG GCG  
 ala his his gly val his gln leu arg ser val pro pro ala arg gly arg arg arg arg  
 101/41 151/51  
 GGA TGT GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG  
 arg lys phe arg phe gly gly ala ala ala AMB leu arg pro pro pro pro arg gly val  
 161/61 211 21  
 GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG

SEQ ID N° 14A

FIGURE 14A (REPLACEMENT FIGURE)

50/185

32/11  
 AAG CCC GGC CGC GAC TGT TTG CCG TTT TGG GGC TCC TAC CAG AAC ACC ACC TGG CGG CCG  
 lys pro gly arg asp cys leu pro phe trp gly ser tyr gln asn thr thr trp arg pro  
 62/21  
 CGC ACC ATG GTG TGC ACC AGT TGC GAT CCG TTC CTC CCG CGC GCG GGC GGC GAC GAC GTC  
 arg thr met val cys thr ser cys asp arg phe leu pro arg ala gly gly asp asp val  
 122/41  
 GAT GCC CGC GCC CCG GCG GCG CAG CTG CGT AGC TCG ACC CGG TCG ACG ACG ACG GGG TCG  
 asp ala arg ala pro ala ala gln leu arg ser ser thr arg ser thr thr thr gly ser  
 182/61  
 GCG GAC CAG TCG GCG ATG TCG AGG CGA TGG CAA TAC AGC GCC TTG GTG CGC GGC CAC ACG  
 ala asp gln ser ala met ser arg arg trp gln tyr ser ala leu val arg gly his thr  
 242/81  
 TCT GAG GTG GCG AAG ACC AGT CCC GCG CCC ACC GGC AGC CGG ATC CGG ATA CGC GGT AC  
 ser glu val ala lys thr ser pro ala pro thr gly ser arg ile arg ile arg gly

SEQ ID N° 15B

FIGURE 15B

33/11  
 AGC CCG GCC GCG ACT GTT TGC CGT TTT GGG GCT CCT ACC AGA ACA CCA CCT GGC GGC CGC  
 ser pro ala ala thr val cys arg phe gly ala pro thr arg thr pro pro gly gly arg  
 63/21  
 GCA CCA TGG TGT GCA CCA GTT GCG ATC GST TCC TCC CGC CGC CGG GCG GCG ACG ACG TCG  
 ala pro trp cys ala pro val ala ile gly ser ser arg ala arg ala ala thr thr ser  
 123/41  
 ATG CCG GAG CGC GCG CCG CGC ACC TCG GTA GGT CGA CGC GGT CGA CGA CGA CGG GGT CGG  
 met pro ala pro arg arg arg ser cys val ala arg pro gly arg arg arg arg gly arg  
 183/61  
 CGG ACC AGT CGG CGA TGT CGA GCG GAT GCG AAT ACA GAG GGT TGG TCG CGC GCG ACA GGT  
 arg thr ser arg arg cys arg gly asp gly arg thr ala pro trp cys ala ala thr arg  
 243/81  
 TCG ACC TCG CGA AGA CGA GGT CCG CGC CGA CCG CGA GAG CGA TCG TCA TAT CGT GAA  
 arg thr thr arg arg arg val pro arg pro pro ala ala arg ser gly tyr ile val

SEQ ID N° 15C

FIGURE 15C

51/185

partie de la séquence nucléotidique de seq15A

```

1/1                               31/11
GGC GGC CGC GCG CCA TGG TGT GCA CCA GTT GCG ATC GGT TCT CCC GCG CGC GGG CGG CGA
gly gly arg ala pro trp cys ala pro val ala ile gly ser pro ala arg gly arg arg
61/21                               91/31
CGA CGT CGA TGG CCG CGC CCC GGC GGC TGC AGC TGC GTA GCT CGA CCC GGT CGA CGA CGA
arg arg arg trp pro arg pro gly gly cys ser cys val ala arg pro gly arg arg arg
121/41                               151/51
CGG GGT CGG CGG GCC AGT CGG CGA TGT CGA GGC GAT GGC AAT ACA GCG CCT TGG TGC GCG
arg gly arg arg ala ser arg arg cys arg gly asp gly asn thr ala pro trp cys ala
181/61                               211/71
GCC ACA CGT CTG AGG TGG CGA AGA CCA GTC GCG CGC CCA CCG GCA GCC GGA TC
ala thr arg leu arg trp arg arg pro val pro arg pro pro ala ala gly

```

SEQ ID N° 15A'

FIGURE 15A'

```

1/1                               31/11
CGG GCC GCG CGC CAT GGT GTG CAC CAG TTG CGA TCG GTT CTC GCG CGC GCG GGC GGC GAC
ala ala ala arg his gly val his gln leu arg ser val leu pro arg ala gly gly asp
61/21                               91/31
GAC GTC GAT GGC CGC GCG CCG GCG GGT GCA GGT GCG TAG CTC GAC CCG GTC GAC GAC GAC
asp val asp gly arg ala pro ala ala ala ala ala AMB leu asp pro val asp asp asp
121/41                               151/51
GGG GTC GCG GGG CCA GTC GCG GAT GTC GAG GCG ATC GCA ATA GAG GCG GTT GGT GCG GCG
gly val gly gly pro val gly asp val ala ala arg ala ile ala asp leu gly ala arg
181/61                               211/71
CCA CAC GTC TGA GGT GTC GAA GAC CAG TCG GCG GCG GAG GTC GAG GCG GAT C
pro his val thr gly gly ala asp gln ser arg ala ala ala ala pro asp

```

SEQ ID N° 15B'

FIGURE 15B'

```

1/1                               31/11
TTC CAC GCG GCG GGT ATG GTC TCG ATT AAT GTC GCG GCG GCG GCG GCG GCG GCG GCG
arg his arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg
61/21                               91/31
ATG GTC GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG
arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg
121/41                               151/51
TTC GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG
arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg
181/61                               211/71
TTC GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG
arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg

```

SEQ ID N° 15C'

FIGURE 15C' (MUTAGENESIS) (continued)

52/185

ORF contenant Seq15A' d'après Cole et al. (Nature 393:537-544)

```

1/1                                31/11
taa ggt ccg cca acg ctt tac gct cga cgg ccg cca cga gtt ggc cgg cca ctt tca ggc
OCH gly pro pro thr leu tyr ala arg arg pro pro arg val gly arg pro leu ser gly
61/21                                91/31
cgt agt cgc cgc agg gca ggg ctt ccc gcg tgg tct tgg cgg gtt tgt cgg caa agg tgt
arg ser arg arg arg ala gly leu pro ala ser ser ser arg val cys arg gln arg cys
121/41                                151/51
agg ggt agc gtt cgt ggg cgt cga cga cga tgt gca gct cgg gga tgc cgg cgg cgc ggg
arg gly ser val arg gly arg arg arg arg cys ala ala arg gly cys arg arg arg gly
181/61                                211/71
cgg tgg ggg tgc gca cgc ccg gcc gcg act gtt tgc gcg ttt tgg ggc tct gcc aga aca
arg trp gly cys ala arg pro ala ala thr val cys ala phe trp gly ser ala arg thr
241/81                                271/91
caa cct ggc ggc cgc gcg cca tgg tgt gca cca gtt gcg atc ggt tct ccc gcg cgc ggg
pro pro gly gly arg ala pro trp cys ala pro val ala ile gly ser pro ala arg gly
301/101                                331/111
cgg cga cga cgt cga tgg ccg cgc ccc gcc ggc tgc agc tgc gta gct cga ccc ggt cga
arg arg arg arg arg trp pro arg pro gly gly cys ser cys val ala arg pro gly arg
361/121                                391/131
cga cga cgg ggt cgg cgg gcc agt cgg cga tgt cga gcc gat gcc aat aca gcg cct tgg
arg arg arg gly arg arg ala ser arg arg cys arg gly asp gly asn thr ala pro trp
421/141                                451/151
tgc gcg gcc aca cgt ctg agg tgg cga aga cca gtc ccg cgc cca ccg gca gcc gga tca
cys ala ala thr arg leu arg trp arg arg pro val pro arg pro pro ala ala gly ser
481/161                                511/171
ggt agg gca gcc gcg agt ctt cag ccg ggt tgg cga cga cga gca gct cca cag agt gtc
gly arg ala gly ala ser leu gln arg gly trp arg arg arg ala ala pro gln ser val
541/181                                571/191
agg gta cgg gcg gcg tcc gcc aac ggt gaa tta tgc aat ccg arg aac cca tgg tca cgt
arg val arg ala ala tyr gly asn gly gln ala gly thr pro thr asn pro ser ser arg
601/201
cga agg gac agg tga
arg arg gly arg OPA

```

SEQ ID NO: 15F

FIGURE 15F

53/185

R:Rv2530c prédite d'après Cole et al. (Nature 393:537-544) et pouvant être en phase avec SEQ15A

```

1/1                                     31/11
gtg acg gca ctg ctc gat gtc aat gtg ctg atc gcg ctg ggc tgg ccg aat cac gtt cac
val thr ala leu leu asp val asn val leu ile ala leu gly trp pro asn his val his
61/21                                     91/31
cat gcg gcc gcg cag cga tgg ttc acg cag ttc tcc tcg aat ggg tgg gcc acc acg ccg
his ala ala ala gln arg trp phe thr gln phe ser ser asn gly trp ala thr thr pro
121/41                                     151/51
atc acc gag gca ggg tat gtc cga att tca agc aat cgc agt gtg atg cag gtg tgg acc
ile thr glu ala gly tyr val arg ile ser ser asn arg ser val met gln val ser thr
181/61                                     211/71
acg ccg gct atc gcg atc gct cag ttg gcg gcg atg act tct ctt gcc ggg cac acg ttt
thr pro ala ile ala ile ala gln leu ala ala met thr ser leu ala gly his thr phe
241/81                                     271/91
tgg cct gac gat gtg cca ctg atc gtt ggg agc gcc ggc gat cgc gat gcg gtg tcc aac
trp pro asp asp val pro leu ile val gly ser ala gly asp arg asp ala val ser asn
301/101                                    331/111
cac cgt cgg gtc acc gag tgc cat ctc atc gcc ttg gcc gcg cgc tac ggg gcc cgg ttg
his arg arg val thr asp cys his leu ile ala leu ala ala arg tyr gly gly arg leu
361/121                                    391/131
gtc aca ttc gat gcc gca ctg gcc gat tca gca tcc gca gcc ctc gtc gag gtg ttg tag
val thr phe asp ala ala leu ala asp ser ala ser ala gly leu val glu val leu AMB

```

SEQ ID N° 15R

FIGURE 15R

Seq15P: ORF d'après Cole et al. (Nature 393:537-544) contenant Rv2530c

```

1/1                                     31/11
tga tgt tcc gcc gga tgc gcc gac ggt gac ttc cga gaa tgt cgt ccg cgc gat cga gga
OPA cys ser ala gly cys ala asp gly asp phe arg gly cys arg pro arg ala arg gly
61/21                                     91/31
cga cgt ggg acc gaa ctg ctc cat ctc aat atg ctc atc gag atg ggt tgg ccg aat cgt
arg arg val thr ala leu leu asp val asn val leu ile ala leu gly trp pro asn his
121/41                                     151/51
ggt cag cat gag gcc agt agt agt ttt ttt agt cgt ttc ttt ttt ttt ttt ttt ttt ttt
val his his ala ala ala ala arg trp phe thr gln phe ser ser asn gly trp ala thr
181/61                                     211/71
atg ttt atg ttt gag ttt ggt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt
thr pro ile thr ala ala gly tyr val arg arg ser ser asn arg ser val met ala val
241/81                                     271/91
tgg aat agc cgt gat atg gca atc ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt
ser thr thr pro ala ala ala ala ala gln leu ala ala met thr ser leu ala gly his
301/101                                    331/111
arg ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt
thr pro thr pro thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr
361/121                                    391/131
tgg aat agc cgt gat atg gca atc ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt
ser thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr
421/151
tgg aat agc cgt gat atg gca atc ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt
ser thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

```

SEQ ID N° 15P

AMBIEN, 1998/01/10

54/185

31/1  
tga tgt tcc gcc gga tgc gcc gac ggt gac ttc cga gga tgt cgt ccg cgc gct cga gga  
OPA cys ser ala gly cys ala asp gly asp phe arg gly cys arg pro arg ala arg gly  
asp val pro pro asp ala pro thr val thr ser glu asp val val arg ala leu glu asp  
Met phe arg arg met arg arg arg OPA leu pro arg met ser ser ala arg ser arg thr  
61/21 91/31  
cga cgt gtg acg gca ctg ctc gat gtc aat gtg ctg atc gcg ctg gcc tgg ccg aat cac  
arg arg val thr ala leu leu asp val asn val leu ile ala leu gly trp pro asn his  
asp val OPA arg his cys ser met ser met cys OPA ser arg trp ala gly arg ile thr  
chr cys asp gly thr ala arg cys gln cys ala asp arg ala gly leu ala glu ser arg  
121/41 151/51  
gtt cac cat gcg gcc gcg cag cga tgg ttc acg cag ttc tcc tcg aat ggg tgg gcc acc  
val his his ala ala ala gln arg trp phe thr gln phe ser ser asn gly trp ala thr  
phe thr met arg pro arg ser asp gly ser arg ser ser pro arg met gly gly pro pro  
ser pro cys gly arg ala ala met val his ala val leu leu glu trp val gly his his  
181/61 211/71  
acg ccg atc acc gag gca ggg tat gtc cga att tca agc aat cgc agt gtg atg cag gtg  
thr pro ile thr glu ala gly tyr val arg ile ser ser asn arg ser val met gln val  
arg arg ser pro arg gln gly met ser glu phe gln ala ile ala val OPA cys arg cys  
ala asp his arg gly arg val cys pro asn phe lys gln ser gln cys asp ala gly val  
241/81 271/91  
tgc acc acg ccg gct atc gcg atc gct cag ttg gcg gcg atg act tet ctt gcc ggg cac  
ser thr thr pro ala ile ala gln leu ala ala met thr ser leu ala gly his  
arg pro arg arg leu ser arg ser leu ser trp arg arg OPA leu leu leu pro gly thr  
asp his ala gly tyr arg asp arg ser val gly gly asp asp phe ser cys arg ala his  
301/101 331/111  
acg ttt tgg cct gac gat gtg cca ctg atc gtt ggg agc gcc gcc gat cgc gat gcg gtg  
thr phe trp pro asp asp val pro leu ile val gly ser ala gly asp arg asp ala val  
arg phe gly leu thr met cys his OPA ser leu gly ala pro ala ile ala met arg cys  
val leu ala OPA arg cys ala thr asp arg trp glu arg arg arg ser arg cys gly val  
361/121 391/131  
tcc aac cad cgt cgg gtc acc gac tgc cat ctc atc gcc ttg gcc gcg cgs tac ggg gcc  
ser asn his arg arg val thr asp cys his leu ile ala leu ala ala arg tyr gly gly  
pro thr thr val gly ser pro thr ala ile ser ser pro trp pro arg ala thr gly ala  
glu pro pro ser gly his arg leu pro ser his arg leu gly arg ala leu arg gly pro  
421/141 451/151  
cag ttg tcc ata ttc gat gcc gca ctg gcc gat tca gca tcc gca gcc ctc gtc gag gtg  
arg leu val thr phe asp ala ala leu ala asp ser ala ser ala gly leu val glu val  
gly trp ser his ser met pro his trp pro ile gln his pro gln ala ser ser arg cys  
val gly his ile asp cys arg thr gly arg phe ser ile arg arg pro arg arg gly val  
481/161 511/171  
tgg tag taa gag gaa aaa gaa gaa tgg tta gga tgg cag gat tta cgg arg cad gcg gaa  
leu AMB ser pro gly met gly gly ser pro gly leu din asp leu arg ala din ala pro  
gaa ser his aa gly trp ala ala arg din ala cys arg ala cys gly arg arg arg pro  
val val thr gly arg gly arg leu ala arg pro ala gly ser ala gly ala gly ala pro  
541/181 571/191  
cag gta gga tcc cga gaa tta gaa gaa ttt ggc cca cga gaa gaa tta gaa gaa gaa gaa  
pro val gly his arg ala ala asp ala phe gly pro arg ala gln leu gly ala ala gly  
arg ser asp thr gly arg pro thr leu leu ala his ala arg ser ser ala leu leu gly  
gly arg thr pro ala gly arg arg phe trp pro thr arg ala ala arg arg cys trp ala  
601/201 631/211  
cag gta gga tcc cga gaa tta gaa gaa ttt ggc cca cga gaa gaa tta gaa gaa gaa gaa  
pro val gly his arg ala ala asp ala phe gly pro arg ala gln leu gly ala ala gly  
arg ser asp thr gly arg pro thr leu leu ala his ala arg ser ser ala leu leu gly  
gly arg thr pro ala gly arg arg phe trp pro thr arg ala ala arg arg cys trp ala  
661/221 691/231  
cag gta gga tcc cga gaa tta gaa gaa ttt ggc cca cga gaa gaa tta gaa gaa gaa gaa  
pro val gly his arg ala ala asp ala phe gly pro arg ala gln leu gly ala ala gly  
arg ser asp thr gly arg pro thr leu leu ala his ala arg ser ser ala leu leu gly  
gly arg thr pro ala gly arg arg phe trp pro thr arg ala ala arg arg cys trp ala

55/185

[illegible]

56/185

1321/441 1351/451  
 ctt ggt gcg cgg cca cac gtc tga ggt ggc gaa gac cag tcc cgc gcc cac cgg cag ccg  
 leu gly ala arg pro his val OPA gly gly glu asp gln ser arg ala his arg gln pro  
 leu val arg gly his thr ser glu val ala lys thr ser pro ala pro thr gly ser arg  
 trp cys ala ala thr arg leu arg trp arg arg pro val pro arg pro pro ala ala gly  
 1381/461 1411/471  
 gat cag gta ggg cag gcg cga gtc ttc agc ggg gtt ggc gcc gac gag cag ctc cac aga  
 asp gln val gly gln ala arg val phe ser gly val gly gly asp glu gln leu his arg  
 ile arg AMB gly arg arg glu ser ser ala gly leu ala ala thr ser ser ser thr glu  
 ser gly arg ala gly ala ser leu gln arg gly trp arg arg arg ala ala pro gln ser  
 1441/481 1471/491  
 gtg tga ggg tac ggg cgg cgt acg gca acg gtg aag cag gca ctc cga cga acc cat cgt  
 val OPA gly tyr gly arg arg thr ala thr val lys gln ala leu arg arg thr his arg  
 cys glu gly thr gly gly val arg gln arg OPA ser arg his ser asp glu pro ile val  
 val arg val arg ala ala tyr gly asn gly glu ala gly thr pro thr asn pro ser ser  
 1501/501  
 cac gtc gaa ggg gca ggt ga  
 his val glu gly ala gly  
 thr ser lys gly gln val  
 arg arg arg gly arg OPA

SEQ ID N° 15Q (suite 2)

FIGURE 15Q (suite (2))

31/11  
 TGC GCA TGC CGA CCA CTC TGG TTG GCC GGA GTT CTT TTT TTT GCG ATT GGC TCA AAG ATT  
 cys ala cys arg pro val trp leu ala gly val arg leu pro ala ile ala ser thr ile  
 61/21 91/31  
 CGA TAT AAC CAC TCT AGT CAG ATC AAC GAT AAT CTT ACC ATT GAG CCG GTG GGT TCA TCC  
 arg tyr asn his ser ser his ile asn his thr ala thr ile glu arg val gly ser cys  
 121/41 151/51  
 GAT GCA TTC GCG ACC GCG GGA GCC GGC GAA TTT GTT TTT AAA CAT AAT CTA GAT TGA GGA  
 his ala phe ala thr ala gly ala gly ala pro gly ala thr his asn pro asp OPA gly  
 181/61 211/71  
 GAG TTT CTT GGT GGA GGA AAA AAA GAT AAT AA TTT TTT GGA AAT CTT GAT GAT GGT TGC GAT  
 asp phe arg ala ala pro thr pro thr ala phe arg ser his ile arg gly arg arg  
 241/81 271/91  
 ... GAT ACC TGT AAT TTT TTT TTT ATA TTT TTT TTT TTA TTA ATT AAT TTA AA GAT TTA  
 ... GAT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT  
 301/101 331/111  
 GAT GAT TTA GCA GCA TCA AAT GAA TTA ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 361/121 391/131  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 421/141 451/151  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 481/161 511/171  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 541/181 571/191  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 601/201 631/211  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 661/221 691/231  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 721/241 751/251  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 781/261 811/271  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 841/281 871/291  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 901/301 931/311  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 961/321 991/331  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 1021/341 1051/351  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 1081/361 1111/371  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 1141/381 1171/391  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 1201/401 1231/411  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 1261/421 1291/431  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 1321/441 1351/451  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 1381/461 1411/471  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 1441/481 1471/491  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 1501/501  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 ... GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT



57/185

32/11  
 GCG CAT GCC GAC CAG TGT GGT TGG CCG GAG TTC GTT TGT TCG CGA TTG CCT CAA CGA TTC  
 ala his ala asp gln cys gly trp pro glu phe val cys ser arg leu pro gln arg phe  
 62/21  
 GAT ATA ACC ACT CTA GTC ACA TCA ACC ACA CTC GTA CCA TCG AGC GTG TGG GTT CAT GCC  
 asp ile thr thr leu val thr ser thr thr leu val pro ser ser val trp val his ala  
 122/41  
 ATG CAT TCG CGA CCG CGG GAG CCG GCG AAC CCG GCG CCA CAC ATA ATC CAG ATT GAG GAG  
 met his ser arg pro arg glu pro ala asn pro ala pro his ile ile gln ile glu glu  
 182/61  
 ACT TCC GTG CCG AAC CGA CGC CGA CGC AAG CTT TCG ACA GCC ATG AGC GCG GTC GCC GCC  
 thr ser val pro asn arg arg arg lys leu ser thr ala met ser ala val ala ala  
 242/81  
 CTG GCA GTT GGA AGT CCT TGT GCA TAT TTT CTT GTC TAC GAA TCA ACC GAA ACG ACC GAG  
 leu ala val ala ser pro cys ala tyr phe leu val tyr glu ser thr glu thr thr glu  
 302/101  
 CCG CCC GAG CAC CAT GAA TTC AAG CAG GCG GCG GTG TTG ACC GAC CTG CCC GCG GAG CTG  
 arg pro glu his his glu phe lys gln ala ala val leu thr asp leu pro gly glu leu  
 362/121  
 ATG TCC GCG CTA TCG CAG GGG TTG TCC CAG TCC GGG ATC  
 met ser ala leu ser gln gly leu ser gln phe gly ile

SEQ ID N° 16B

FIGURE 16B

33/11  
 CGC ATG CCG AGC AGT GTG GTT GCG CCG AGT TCG TTT GTT CCG GAT TCG CTC AAC GAT TCG  
 arg met pro thr ser val val gly arg ser ser phe val arg asp cys leu asn asp ser  
 63/21  
 ATA TAA CCA CTC TAG TCA CAT CAA CCA CAC TCG TAC CAT CGA GCG TGT GCG TTC ATG CCA  
 ile och pro leu asp ser his gln pro his ser tyr his ala ala tyr gly phe met pr  
 123/41  
 TTT ATT CCG GAT GCG GCG AGC GCG CCA AAC CCG CCG CAC AAA TAA TCG AGA TTC ATT AAA  
 ty ile arg arg arg gly ser his arg thr arg arg his thr thr ser arg leu arg arg  
 183/61  
 GTT CCG TTC CGA ACC GAT GCG GAT GGA ATT TTT TCA CAG CCA TTA GAT GAT GAT GAT  
 leu le cys arg thr asp ala arg ala ser phe arg thr pr CTA phe arg thr pr pr  
 243/81  
 TTA CAG TTC CAA TTC TTT GTG AT ATT TTT TTG TGT ATG AAT CAA CCG AAA TCA TCA AAT  
 thr gla leu ala val leu val hi ile phe leu ser thr arg thr pr CTA arg pr leu  
 303/101  
 TTT TTT AAT TTT AAT TTT TTT AAT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT  
 ty ty thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
 363/121

FIGURE 16C

SEQUENCE ALIGNMENT

58/185

31/11  
 GCG GGC CAC CGA TCA GTC GAT CGG GTG GTT TCC GCT CCA TCA GCC CGG AAT TGA GGT GCC  
 ala gly his arg ser val asp arg val val ser ala pro ser ala arg asn OPA gly ala  
 61/21  
 GCA GTG ACG ACA CCA GCG CAG GAC GCG CCG TTG GTG TTT CCC TCT GTT GCT TTC CCG TCC  
 ala val thr thr pro ala gln asp ala pro leu val phe pro ser val ala phe pro ser  
 121/41  
 GGC TCG CCT TTT TTT CAT CAA CGT TGG ACT GCC GCA GTG GCG ATG TTG GTC GCC GGC GTG  
 gly ser pro phe phe his gln arg trp thr ala ala val ala met leu val ala gly val  
 181/61  
 TTC GGT CAC CTG ACG GTC GGG ATG TTC CTT GGG TCT CCG GTT GCT GCT GGG TTT GCT CAA  
 phe gly his leu thr val gly met phe leu gly ser arg val ala ala gly phe ala gln  
 241/81  
 TGC CCT GCT GGT GCG GCG TTC GGC CGA GTC GAT CAC CGC CAA AGA GCA CCC GTT AAA ACG  
 cys pro ala gly ala ala phe gly arg val asp his arg gln arg ala pro val lys thr  
 301/101  
 GTC GAT GGC CCT CAA CTC GGC ATC GCG ACT GGC GAT TAT CAC CAT GCC TGG GGC TGA TC  
 val asp gly pro gln leu gly ile ala thr gly asp tyr his his ala ser gly OPA

SEQ ID N° 17A

FIGURE 17A

32/11  
 CGG GCC ACC GAT CAG TCG ATC GGG TGG TTT GCG CTC GAT CAG CCC GGA ATT GAG GTG CCG  
 arg ala thr asp gln ser ile gly trp phe pro leu his ala pro gly ile gln val pro  
 62/21  
 CAG TGA CGA CAC CAG CCG AGG ACG CGC CTT TGG TGT TTT CCT CTC TTG CTT TCC CCG CCG  
 gln OPA arg his gln arg arg thr arg arg trp cys phe pro leu leu leu ser arg pro  
 122/41  
 GGT CGC CTT TTT TTC ATC AAC GTT GGA CTG CCG CAG TGG CGA TGT TGG TGG CCG GCG TGT  
 ala arg leu phe phe ile asn val gly leu thr gln trp gln cys trp asp pro ala cys  
 182/61  
 TGG CTC AAT TGA TGG TTT GGA TAT TTT TTT GAT CTC GCG TTT CTC CTC GGT TTT CTC AAT  
 ser val thr OPA arg ser gly cys ser leu gly leu leu leu gly leu leu leu  
 242/81  
 TTT CTC CTC CTC TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT  
 ala leu leu val arg arg ser ala ala ser ile thr ala lys gln his pro leu lys arg  
 302/101  
 TGG ATG GTC CTC AAG TTT CGA TTT CGA CTC CCG AAT AAT AAT AAT AAT AAT AAT AAT AAT  
 leu met ala leu arg leu ala ser arg leu ala ala ala ala ala ala ala ala ala ala

SEQ ID N° 17B

59/185

33/11  
 GGG CCA CCG ATC AGT CGA TCG GGT GGT TTC CGC TCC ATC AGC CCG GAA TTG AGG TGC CGC  
 gly pro pro ile ser arg ser gly gly phe arg ser ile ser pro glu leu arg cys arg  
 63/21  
 AGT GAC GAC ACC AGC GCA GGA CGC GCC GTT GGT GTT TCC CTC TGT TGC TTT CCC GTC CGG  
 ser asp asp thr ser ala gly arg ala val gly val ser leu cys cys phe pro val arg  
 123/41  
 CTC GCC TTT TTT TCA TCA ACG TTG GAC TGC CGC AGT GGC GAT GTT GGT CGC CGG CGT GTT  
 leu ala phe phe ser ser thr leu asp cys arg ser gly asp val gly arg arg arg val  
 183/61  
 CGG TCA CCT GAC GGT CGG GAT GTT CCT TGG GTC TGG GGT TGC TGC TGG GTT TGC TCA ATC  
 arg ser pro asp gly arg asp val pro trp val ser gly cys cys trp val cys ser met  
 243/81  
 CCC TGC TGG TGC GGC GTT CGG CCG AGT CGA TCA CCG CCA AAG AGC ACC CCT TAA AAG GGT  
 pro cys trp cys gly val arg pro ser arg ser pro pro lys ser thr arg OCH asn gly  
 303/101  
 CGA TGG CCC TCA ACT CGG CAT CGC GAC TGG CGA TTA TCA CCA TGC CTC GGG CTG ATC  
 arg trp pro ser thr arg his arg asp trp arg leu ser pro cys leu gly leu ile

SEQ ID N° 17C

FIGURE 17C

partie de la séquence nucléotidique de se17A

1/1  
 31/11  
 ggg tag aac ccc gaa gga gac ctc gag ggt ttc ccc tcc atc agc ccc gaa ttg agg tgc cgc  
 gly AMB asn pro glu gly asp leu ala gly cys arg arg pro ala his arg met arg ile  
 61/21  
 91/41  
 cgg tgg cga cga tta asg acc gac ata ggg att tta cct tta agt gat tcc ggt gag acc  
 arg ser arg arg phe thr thr asp ile gly ser trp pro leu gly asp ser gly ala thr  
 121/41  
 141/61  
 act gag ata cga tgg gag gga cac cga tca gtc tta att gta gtt tcc ggt cca tca ggc  
 thr ala ile asp ser ala gly his arg ser val val val val val val val val val val  
 161/81  
 181/101  
 cgg aat tca ggt tca ggt gga gga gga gga gga gga gga gga gga gga gga gga gga gga  
 arg asp tca ggt tca ggt gga gga gga gga gga gga gga gga gga gga gga gga gga gga  
 201/121  
 221/141  
 gtt ggt tta ggt tta ggt tta ggt tta ggt tta ggt tta ggt tta ggt tta ggt tta ggt tta  
 val ala phe arg pro val arg leu phe phe ile arg val gly leu ala ala val ala val  
 241/161  
 261/181  
 tgg  
 leu val ala gly val phe gly thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
 281/201  
 301/221  
 tgg  
 leu gly thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

FIG. 17C

SEQUENCE N° 17C

60/185

```

1/1                               31/11
gct aga acc ccg aag gag acc tcg cgg gtt gcc gcc ccc cgg ccc atc gga tgc gta tcc
ala arg thr pro lys glu thr ser arg val ala gly pro arg pro ile gly cys val ser
61/21                               91/31
ggt cgc gcc gat tca cga ccg aca tag gga gct acc cct tgg gtg att ccg gtg cga cga
gly arg ala asp ser arg pro thr AMB gly ala thr pro trp val ile pro val arg arg
121/41                               151/51
ctg cga tac gct cgg cgg gcc acc gat cag tcg atc ggg tgg ttt ccg ctc cat cag ccc
leu arg tyr ala arg arg ala thr asp gln ser ile gly trp phe pro leu his gln pro
181/61                               211/71
gga att gag gtg ccg cag tga cga cac cag cgc agg acg cgc cgt tgg tgt ttc cct ctg
gly ile glu val pro gln OPA arg his gln arg arg thr arg arg trp cys phe pro leu
241/81                               271/91
ttg ctt tcc gtc cgg ttc gcc ttt ttt tca tca acg ttg gac tgg ccg cag tgg cga tgt
leu leu ser val arg phe ala phe phe ser ser thr leu asp trp pro gln trp arg cys
301/101                               331/111
tgg tcg ccg gcc tat tcg gtc acc tga cgg tcg gga tgt tct tgg gtc tgg ggt tgc tgc
trp ser pro ala cys ser val thr OPA arg ser gly cys ser trp val ser gly cys cys
361/121                               391/131
tgg gtt tgc tca atg ccc tgc tgg tgc gcc gtt cgg ccg agt cga tca ccg cca aag agc
trp val cys ser met pro cys trp cys gly val arg pro ser arg ser pro pro lys ser
421/141                               451/151
acc cgt taa aac ggt cga tgg ccc tca act cgg cat cgc gac tgg cga tta tca cca tcc
thr arg OCH asn gly arg trp pro ser thr arg his arg asp trp arg leu ser pro ser
481/161
tcg gcc tga tc
ser gly OPA

```

SEQ ID N° 17B'

FIGURE 17B'

```

1/1                               31/11
cta gaa ccc cna agg aga cct cgg ggg ttg cgg gcc ccc cgg ccc atc gga tgc gta tcc
leu glu pro arg arg arg pro arg alv leu pro ala gaa gly pro ser asp ala tyr pro
61/21                               91/31
gtc gcc ccg att cag gat cga cat acc gag cta ccc att gaa taa ttc cgg tgc gaa gat
val ala pro ile his asp arg his arg glu leu pro leu gly OPA phe arg cys asp asp
121/41                               151/51
tgg gat arg tc tgg ggg cca cca tta arg cna tca cgt gat tta cca tca atc cca cca
cys asp thr tc alv gly pro pro ile ser and ser gly arg pro arg met ile ser pr
181/61                               211/71
tta tta acc tta tta att acc tta tta tta tta tta tta tta tta tta tta tta tta tta
tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta
241/81                               271/91
tga tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta
cys phe pro ser gly ser pro phe phe phe phe phe phe phe phe phe phe phe phe phe
301/101                               331/111
ttr cgg cgg ttr acc cgg ttr ttr acc ttr ttr acc ttr ttr acc ttr ttr acc ttr ttr acc
ty arg arg arg val arg ser pro arg ty arg arg val arg ty arg arg val arg ty arg
361/121                               391/131
tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta
ty phe tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta
421/141                               451/151
tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta
ty phe tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta

```

SEQ ID N° 17B'

FIGURE 17B'

61/185

séquence Rv1303 prédite par Cole et al. (Nature 393:537-544) et contenant partiellement Seq17A'

```

1/1                                     31/11
atg acg aca cca gcg cag gac gcg ccg ttg gtg ttt ccc tct gtt gct ttc cgt ccg gtt
met thr thr pro ala gln asp ala pro leu val phe pro ser val ala phe arg pro val
61/21                                     91/31
cgc ctt ttt ttc atc aac gtt gga ctg gcc gca gtg gcg atg ttg gtc gcc gcc gtg ttc
arg leu phe phe ile asn val gly leu ala ala val ala met leu val ala gly val phe
121/41                                     151/51
ggt cac ctg acg gtc ggg atg ttc ttg ggt ctc ggg ttg ctg ctg ggt ttg ctc aat gcc
gly his leu thr val gly met phe leu gly leu gly leu leu leu gly leu leu asn ala
181/61                                     211/71
ctg ctg gtg ccg cgt tgc gcc gag tgc atc acc gcc aaa gag cac ccg tta aaa ccg tgc
leu leu val arg arg ser ala glu ser ile thr ala lys glu his pro leu lys arg ser
241/81                                     271/91
atg gcc ctc aac tgc gca tgc cga ctg gcg att atc acc atc ctc ggg ctg atc atc gcc
met ala leu asn ser ala ser arg leu ala ile ile thr ile leu gly leu ile ile ala
301/101                                    331/111
tac att ttc ccg ccc gct gga ttg gcc gtc gtg ttc ggg ctg gca ttc ttc cag gtg ctg
tyr ile phe arg pro ala gly leu gly val val phe gly leu ala phe phe gln val leu
361/121                                    391/131
ctg gtg gca acg acg gcc ctg ccg gtc ctg aag aag ctg cgc act gcg acc gag gaa ccg
leu val ala thr thr ala leu pro val leu lys lys leu arg thr ala thr glu glu pro
421/141                                    451/151
gtc gca act tat tct tcc aat gcc cag acc ggg gga tgc gaa gga agg agc gcc agc gat
val ala thr tyr ser ser asn gly gln thr gly gly ser glu gly arg ser ala ser asp
481/161
gac tga
asp OPA

```

SEQ ID N° 17D

FIGURE 17D

Orf d'après Cole et al. (Nature 393:537-544) et contenant Rv1303

```

1/1                                     31/11
tga ggt gcc gca gtc acg aca cca gcg cag gac gcg ccg ttg gtg ttt ccc tct gtt gct
OPA gly ala ala val thr thr pro ala gln asp ala pro leu val phe pro ser val ala
61/21                                     91/31
ttc cgt ccg gtt cgc ctt ttt ttc atc aac gtt gga ctg gcc gca gtg gcg atg ttg gtc
phe arg pro val arg leu phe phe ile asn val gly leu ala ala val ala met leu val
121/41                                     151/51
gga gcc gtc ttc ggt cac ctg acg gtc ggg atg ttc ttg ggt ctc ggg ttg ctg ctg ggt
ala gly val phe gly his leu thr val gly met phe leu gly leu gly leu leu leu gly
181/61                                     211/71
tac att ttc ccg ccc gct gga ttg gcc gtc gtg ttc ggg ctg gca ttc ttc cag gtg ctg
leu leu asn ala leu leu val arg arg ser ala ile ser ile thr ala lys his arg pro
241/81                                     271/91
tta aaa ccg tgc atg gcc ctc aac tgc gca tgc cga ctg gcg att atc acc atc ctc ggg
leu lys ala ser met ala leu asn ser ala ser arg leu ala ile ile thr ile leu gly
301/101                                    331/111
ctg att ttc ccg ccc gct gga ttg gcc gtc gtg ttc ggg ctg gca ttc ttc cag gtg ctg
leu leu asn ala leu leu val arg arg ser ala ile ser ile thr ala lys his arg pro
361/121                                    391/131
ctg gtg gca acg acg gcc ctg ccg gtc ctg aag aag ctg cgc act gcg acc gag gaa ccg
leu val ala thr thr ala leu pro val leu lys lys leu arg thr ala thr glu glu pro
421/141                                    451/151
gtc gca act tat tct tcc aat gcc cag acc ggg gga tgc gaa gga agg agc gcc agc gat
val ala thr tyr ser ser asn gly gln thr gly gly ser glu gly arg ser ala ser asp
481/161
gac tga
asp OPA

```

SEQ ID N° 17E

EMPLACEMENT (Rv1303)

31/11

FIGURE 18A

32/11

TGG AAG AGG TAG GGA AGG GGG GGT GGG TGG GTC GGT GAG GTC GTA TGT GGT GTT CAG GGA  
ser asn arg cys gly arg arg arg arg ser leu gly pro leu val ser arg val gln pro  
52/21

52/31

GCC AGG GGC CGT TAA GGT GGG GGA AGA GGT GGT GTT GGG GTC GGG CAT CAG GGT GGA TGT  
ala ser gly arg CPH arg gly arg thr gly arg leu gly val gly his ala arg arg tps  
122/41

122/51

GGG TAA GAG TAA TAA TGG AAG GGA TGG GAA GGA TGA TAT GAT GAT TAT AAT TGT TTT  
gly ser gly arg tps tps arg gly thr gln val ser tps arg his pro ser thr ser pro  
122/61

TGG GTG TAA TAA TGG GGT CAT GGT TGA GGT GGA TGA GAT TAA TAT AAT GGG GGA AGA AAA  
arg val arg arg arg arg gly his arg CPH arg gly ala gln pro ala thr gly tps arg arg  
122/81

TGG TGA AAA GAA TT AAG G GGA TGG GGT GGA GGT GAT GGT GGT TGA AAT GAA ATT AAT  
arg gly arg arg leu ala ala arg arg arg gly gly arg arg gly arg thr ala thr tps  
122/101

122/111

122/121

122/131

122/141

122/151

122/161

122/171

122/181

122/191

122/201

122/211

122/221

122/231

122/241

122/251

122/261

122/271

122/281

122/291

122/301

122/311

122/321

122/331

122/341

122/351

122/361

122/371

122/381

122/391

122/401

122/411

122/421

122/431

122/441

122/451

122/461

122/471

122/481

122/491

122/501

122/511

122/521

122/531

122/541

122/551

122/561

122/571

122/581

122/591

122/601

122/611

122/621

122/631

122/641

122/651

122/661

122/671

122/681

122/691

122/701

122/711

122/721

122/731

122/741

122/751

122/761

122/771

122/781

122/791

122/801

122/811

122/821

122/831

122/841

122/851

122/861

122/871

122/881

122/891

122/901

122/911

122/921

122/931

122/941

122/951

122/961

122/971

122/981

122/991

123/1

123/2

123/3

123/4

123/5

123/6

123/7

123/8

123/9

123/10

123/11

123/12

123/13

123/14

123/15

123/16

123/17

123/18

123/19

123/20

123/21

123/22

123/23

123/24

123/25

123/26

123/27

123/28

123/29

123/30

123/31

123/32

123/33

123/34

123/35

123/36

123/37

123/38

123/39

123/40

123/41

123/42

123/43

123/44

123/45

123/46

123/47

123/48

123/49

123/50

123/51

123/52

123/53

123/54

123/55

123/56

123/57

123/58

123/59

123/60

123/61

123/62

123/63

123/64

123/65

123/66

123/67

123/68

123/69

123/70

123/71

123/72

123/73

123/74

123/75

123/76

123/77

123/78

123/79

123/80

123/81

123/82

123/83

123/84

123/85

123/86

123/87

123/88

123/89

123/90

123/91

123/92

123/93

123/94

123/95

123/96

123/97

123/98

123/99

124/1

124/2

124/3

124/4

124/5

124/6

124/7

124/8

124/9

124/10

124/11

124/12

124/13

124/14

124/15

124/16

124/17

124/18

124/19

124/20

124/21

124/22

124/23

124/24

124/25

124/26

124/27

124/28

124/29

124/30

124/31

124/32

124/33

124/34

124/35

124/36

124/37

124/38

124/39

124/40

124/41

124/42

124/43

124/44

124/45

124/46

124/47

124/48

124/49

124/50

124/51

124/52

124/53

124/54

124/55

124/56

124/57

124/58

124/59

124/60

124/61

124/62

124/63

124/64

124/65

124/66

124/67

124/68

124/69

124/70

124/71

124/72

124/73

124/74

124/75

124/76

124/77

124/78

124/79

124/80

124/81

124/82

124/83

124/84

124/85

124/86

124/87

124/88

124/89

124/90

124/91

124/92

124/93

124/94

124/95

124/96

124/97

124/98

124/99

125/1

125/2

125/3

125/4

125/5

125/6

125/7

125/8

125/9

125/10

125/11

125/12

125/13

125/14

125/15

125/16

125/17

125/18

125/19

125/20

125/21

125/22

125/23

125/24

125/25

125/26

125/27

125/28

125/29

125/30

125/31

125/32

125/33

125/34

125/35

125/36

125/37

125/38

125/39

125/40

125/41

125/42

125/4

63/185

3/1 33/11  
 CGA ACA GGT ACG GAA GGC GCC GTC GGT CGC TCG GTC CGC TGG TAT CTC GTG TTC AGC CAG  
 arg thr gly thr glu gly ala val gly arg ser val arg trp tyr leu val phe ser gln  
 63/21 93/31  
 CCA GCG GCC GTT AAC GTG GCC GAA CAG GTC GTC TTG GGG TCG GGC ATC AGC GTC GAT GTG  
 pro ala ala val asn val ala glu gln val val leu gly ser gly ile ser val asp val  
 123/41 153/51  
 GCT CAG GTC GAT ACC CGA GGG GAT GGC AAG TGT CAC CCC GCC ATC CTT CCA CCT CTT TTC  
 ala gln val asp thr arg gly asp gly lys cys his pro ala ile leu pro pro leu phe  
 183/61 213/71  
 GGG TGC AAC GAT CGG GCC ATG CCT SAC GGG GAG CAG AGC CAG CCA CCG GCC CAA GAA GAT  
 gly cys asn asp arg ala met pro asp gly glu gln ser gln pro pro ala gln glu asp  
 243/81 273/91  
 GGG GAA GAC GAC TCG CGG CCC GAC GCC GCG GAG GCG SCC GCG GCC GAA CCC AAA TCA TCA  
 ala glu asp asp ser arg pro asp ala ala glu ala ala ala ala glu pro lys ser ser  
 303/101 333/111  
 GGC GGT CCG GAT GTT CTC GAC CTA CCG TAT GCG GTT GAC ACT ACT CCG CCG GGT ATC GGT  
 ala gly pro asp val leu asp leu arg tyr arg leu asp thr thr arg arg ala ile gly  
 363/121  
 CGC CGC GGT CGT GCT GGG TGC GAT GAT C  
 arg arg gly arg ala gly cys asp asp

SEQ ID N° 18C

FIGURE 18C

partie de la séquence nucléotidique de l'ADN

1/1 11/11  
 GAA GGC GGC GTC GGT GGC TCG GTC GCG TGG TAT CTC GTG TTC AGC CAG CCA GAA GGT GTT  
 glu gly ala val gly arg ser val arg trp tyr leu val phe ser gln pro ala ala val  
 61/21 91/31  
 AAC GTG GCC GAA CAG GTC GTC TTG GGG TCG TGG TAT CAC CCC GCC ATC CTT CCA CCT CTT TTC  
 asn val ala glu gln val val leu gly ser gly ile ser val asp val ala gln val asp  
 121/41 151/51  
 ACC GGA GGA GAT GAT AAC TTT CAC GGT ATC AAT GGT CCA GAT CTT TTC AAT TGG AAC GAT  
 thr arg gly asp gly ty ty his pro ala ala cys pro pro leu phe gly thr ala arg  
 181/61 211/71  
 TCG GCG ATC CTT GAT GTC GAC TAT AAT CAC CAG CCG GAG GAG GAG GAG GAG GAG GAG  
 arg ala met pr asp gly ala gln ser gln pr ala gln ala asp ala gln asp arg  
 241/81 271/91  
 TCG CCG CTT GAT GAT GTC GAC GGT GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG  
 ser arg pr ala ala ala gln ala ala gln ala ala gln ala gln ala gln ala gln ala  
 301/101 331/111  
 TTT TTT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
 thr  
 361/121

FIGURE 18C

FIGURE 18C - SEQUENCE NUCLEOTIDIC

64/185

1/1 31/11  
 CGG AAG GCG CCG TCG GTC GCT CGG TCC GCT GGT ATC TCG TGT TCA GCC AGC CAG CGG CCG  
 arg lys ala pro ser val ala arg ser ala gly ile ser cys ser ala ser gln arg pro  
 61/21 91/31  
 TTA ACG TGG CCG AAC AGG TCG TCT TGG GGT CGG GCA TCA GCG TCG ATG TGG CTC AGG TCG  
 leu thr trp pro asn arg ser ser trp gly arg ala ser ala ser met trp leu arg ser  
 121/41 151/51  
 ATA CCC GAG GGG ATG GCA AGT GTC ACC CCG CCA TCC TTC CAC CTC TTT TCG GGT GCA ACG  
 ile pro glu gly met ala ser val thr pro pro ser phe his leu phe ser gly ala thr  
 181/61 211/71  
 ATC GGG CCA TGC CTG ACG GGG AGC AGA GCG AGC CAC CCG CCG AAG AAG ATG CGG AAG ACG  
 ile gly pro cys leu thr gly ser arg ala ser his arg pro lys lys met arg lys thr  
 241/81 271/91  
 ACT CGC GGC CCG ACG CCG CGG AGG CCG CCG CGG CCG AAC CCA AAT CAT CAG CCG GTC CGA  
 thr arg gly pro thr pro arg arg pro pro arg pro asn pro asn his gln pro val arg  
 301/101 331/111  
 TGT TCT CGA CCT ACG GTA TCG CCT CGA CAC TAC TCG GCG TGG TAT CCG TCG CCG CGG TCG  
 cys ser arg pro thr val ser pro arg his tyr ser ala cys tyr arg ser pro arg ser  
 361/121  
 TGC TGG GTG CGA TGA TC  
 cys trp val arg opa

SEQ ID N° 18B'

FIGURE 18B'

1/1 31/11  
 GGA AGT GAG GGT CGG TCG CTC GGT CCG CTC GTA TGT TGT GTT TAA GAA GGG AGT GGT TGT  
 gly arg arg arg arg ser leu gly pro leu val ser arg val gln pro ala ser gly arg  
 61/21 91/31  
 TAA CGT GCG CCA ACA GGT GGT GTT GGG GTC GGG CAT TAA GGT GGA TGT GGC TCA GGT CGA  
 stop arg gly arg thr cly arg leu gly val cly his ala arg thr tyr cly ser gly arg  
 121/41 151/51  
 TAT CCG AAT GAA TGA CAA ATG TTA CCG CCG CAT TGT TGT AAT TGT TTT CCG GTC TAA TAA  
 tyr his arg gly thr ala val ser pro arg his pro arg thr ser pro arg val ala his  
 181/61 211/71  
 TCG GCG CAT GCG TCA CCG GGA GAA GAT TAA GGT AGT TGT TGT A A A A TTT TTA AAT TAA  
 ser gly his ala gaa arg cly arg ala ser ala thr cly pro thr arg tyr gly arg arg  
 241/81 271/91  
 CTC GAT GCG CCA GGT GGT GGA GAT GAT TGT TGT TAA AAT GAA ATG AAT AAT GGT TGT GAT  
 leu ala ala arg arg arg gly cly arg arg arg arg thr ala ala ala ala ala ala ala  
 301/101 331/111  
 TTT TT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT  
 val  
 361/121

FIGURE 18B'

SEQUENCE LISTING



65/185

séquence Rv0199 prédite par Cole et al. (Nature 393:537-544) et contenant seq18A'

```

1/1                               31/11
atg cct gac ggg gag cag agc cag cca ccg gcc caa gaa gat gcg gaa gac gac tcg cgg
Met pro asp gly glu gln ser gln pro pro ala gln glu asp ala glu asp asp ser arg
61/21                               91/31
ccc gac gcc gcg gag gcc gcc gcg gcc gaa ccc aaa tca tca gcc ggt ccg atg ttc tcg
pro asp ala ala glu ala ala ala ala glu pro lys ser ser ala gly pro met phe ser
121/41                               151/51
acc tac ggt atc gcc tcg aca cta ctc gcc gtg cta tcg gtc gcc gcg gtc gtg ctg ggt
thr tyr gly ile ala ser thr leu leu gly val leu ser val ala ala val val leu gly
181/61                               211/71
gcg atg atc tgg tcc gca cac cgc gat gac tcc gcc gag cgt acc tac ctg acc cgg gtc
ala met ile trp ser ala his arg asp asp ser gly glu arg thr tyr leu thr arg val
241/81                               271/91
atg ctg acc gcc gct gaa tgg acg gcc gtg ctg atc aac arg aac gcc gac aac atc gat
met leu thr ala ala glu trp thr ala val leu ile asn met asn ala asp asn ile asp
301/101                               331/111
gcc agc ctg cag cga ctg cac gac gga acg gtc ggt caa ctc aac acc gac ttc gac gct
ala ser leu gln arg leu his asp gly thr val gly gln leu asn thr asp phe asp ala
361/121                               391/131
gtc gtg cag ccc tac cgg cag gtg gtg gag aag ttg cgg acg cac agc agc gcc agg atc
val val gln pro tyr arg gln val val glu lys leu arg thr his ser ser gly arg ile
421/141                               451/151
gag gcg gta gcg atc gat acg gtg cac cgc gag ctg gat acc cag tcc ggt gcc gcc cga
glu ala val ala ile asp thr val his arg glu leu asp thr gln ser gly ala ala arg
481/161                               511/171
ccg gta gta acc acg aaa ttg cca ccg ttt gcc act cgc acc gac tcg gtg ctg ctg gtc
pro val val thr thr lys leu pro pro phe ala thr arg thr asp ser val leu leu val
541/181                               571/191
gcg acg tcg gtc agt gag aac gcc gcc gcc aaa ccc cag acc gty cac tgg aac ttg cgg
ala thr ser val ser glu asn ala gly ala lys pro gln thr val his trp asn leu arg
601/201                               631/211
ctc gat gtc ttc gat gty gat gcc aag ctg atg atc tcc cgg ttg gat tcg att caa tga
leu asp val ser asp val asp gly lys leu met ile ser arg leu ala ser ile arg CDA

```

SEQ ID N° 18D

FIGURE 18D

66/185

ORF d'après Cole et al. (Nature 393:537-544) et contenant Rv0199

1/1  
 taa tcc gat gcc gga ttg ggt gaa atg cac caa gta acg ggt cga gtc ttt gga atc ggt  
 OCH ser asp ala gly leu gly glu met his gln val thr gly arg val phe gly ile gly  
 61/21  
 atc gac ata gac tcc gat gcc gcc gcc cac gcc gcc acg ttg cag agt gcc aag gcc gcc  
 ile asp ile asp ser asp ala ala ala his ala gly thr leu gln ser ala lys gly gly  
 121/41  
 gcc caa ttc ggt gcc gtc gcc cgc gct gtc aat cgt gcc caa ttc gtc gtg cag cgg ttg  
 gly gln phe gly gly val gly arg ala val asn arg gly gln phe val val gln arg leu  
 181/61  
 cac ccc tgc gcg ctc gac gcc ttc ctc gtc gag gaa gct gcc gta gag gtc gcc gat gcg  
 his pro cys ala leu asp gly phe leu val glu glu ala gly val glu val ala asp ala  
 241/81  
 ctg cgc atc ggt gcc tac cgc agc acc tgc ttg gct gcc ctg gat gat caa gtc tgc cac  
 leu arg ile gly ala tyr arg ser thr cys leu ala gly leu asp asp gln val ser his  
 301/101  
 ttg tgt ctc gcc gcg gtc gaa cag gct acg gaa gcc gcc gtc ggt cgc tgc gtc cgc tgg  
 leu cys leu gly ala val glu gln ala thr glu gly ala val gly arg ser val arg trp  
 361/121  
 tat ctc gtg ttc agc cag cca gcg gcc gtt aac gtg gcc gaa cag gtc gtc ttg ggg tgc  
 tyr leu val phe ser gln pro ala ala val asn val ala glu gln val val leu gly ser  
 421/141  
 gcc atc agc gtc gat gtg gct cag gtc gat acc cga ggg gat gcc aag tgt cac ccc gcc  
 gly ile ser val asp val ala gln val asp thr arg gly asp gly lys cys his pro ala  
 481/161  
 atc ctt cca cct ctt ttc ggg tgc aac gat cgg gcc atg cct gac ggg gag cag agc cag  
 ile leu pro pro leu phe gly cys asn asp arg ala met pro asp gly glu gln ser gln  
 541/181  
 cca cgc gcc caa gaa gat gcg gaa gac gac tgc cgg ccc gcc gcc gcg gag gcc gcc gcg  
 pro pro ala gln glu asp ala glu asp asp ser arg pro asp ala ala glu ala ala ala  
 601/201  
 gcc gaa ccc aaa tca tca gcc ggt ccg atg ttc tgc acc tac ggt atc gcc tgc aca cta  
 ala glu pro lys ser ser ala gly pro met phe ser thr tyr gly ile ala ser thr leu  
 661/221  
 ctc gcc gtg cta tgc gtc gcc gcg ctc gtc ctg ggt gcg atg atc tgg tgc gca cgc cgc  
 leu gly val leu ser val ala ala val val leu gly ala met ile trp ser ala his arg  
 721/241  
 gat gac tcc gcc gag cgt acc tac ctg acc cgt ctc atg atg acc gcc gct gaa tgc acc  
 asp asp ser gly glu arg thr tyr leu thr arg val met leu thr ala ala glu trp thr  
 781/261  
 gcc gtg ctg atc aac atg aac gcc gac aac atc atc gcc asp ctg cgt cgt cgt cgt gct  
 ala val leu ile asp met asp ala asp asp ile asp ala ser leu gln asp leu his asp  
 841/281  
 gta acg atc gct cca ctc aac aac aac ttc acc cgt ctc gln cca cgt ta cgc aac gct  
 gly thr val gly gln ile asp thr asp phe asp ala val val ala met ty arg gln val  
 901/301  
 tta gta aac tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta  
 val ala lys leu arg thr his ser ser gly arg ile ala ala val ala ile asp thr val  
 961/321  
 taa cga gaa ctg gat acc gat ttc gct gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc  
 ala arg ala leu asp thr gln ser gly ala ala arg pro val val thr thr gly leu pr  
 1021/341  
 tta tta gta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta  
 thr thr ala thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
 1081/361

GTT GCG CAA CGG GGT GAG CAC CGA CGC GAT GAT GGC GCA ACT ATC GAA ACT GCA GGA CAT  
 val ala gln arg gly glu his arg arg asp asp gly ala thr ile glu thr ala gly his  
 61/21 91/31  
 CGC CAA CGC CAA CGA CGG CAC TCG CGC GGT GGG CAC CCC TGG CTA TCA GGC CAG CGT CGA  
 arg gln arg gln arg arg his ser arg gly gly his pro trp leu ser gly gln arg arg  
 121/41 151/51  
 CTA TGT GGT AAA CAC ACT GCG CAA CAG CGG TTT TGA TGT GCA AAC CCC GGA GTT CTC CGC  
 leu cys gly lys his thr ala gln gln arg phe opa cys ala asn pro gly val leu arg  
 181/61 211/71  
 TCG CGT GTT CAA GGC CGA AAA AGG GGT GGT GAC CCT CGG CGG CAA CAC CGT GGA GGC GAG  
 ser arg val gln gly arg lys arg gly gly asp pro arg arg gln his arg gly gly glu  
 241/81 271/91  
 GGC GCT CGA GTA CAG CCT CGG CAC ACC GCC GGA CGG GGT GAC GGG CCC CCT GGT GGC TGC  
 gly ala arg val gln pro arg his thr ala gly arg gly asp gly pro ala gly gly cys  
 301/101 331/111  
 CCG CGC CGA CGA CAG TCC GGG CTG CAG TCC GTC GGA CTA GGA CAG GCT GCC GGT GTC CGG  
 pro arg arg arg gln ser gly leu gln ser val gly leu arg gln ala ala gly val arg  
 361/121  
 TGC GGT GGT GCT GGT AGA TC  
 cys gly gly ala gly arg

FIGURE 19A

[illegible]

68/185

33/11  
TGC GCA ACG GGG TGA GCA CCG ACG CGA TGA TGG CGC AAC TAT CGA AAC TGC AGG ACA TCG  
cys ala thr gly OPA ala pro thr arg OPA trp arg asn tyr arg asn cys arg thr ser  
63/21  
CCA ACG CCA ACG ACG GCA CTC GCG CGG TGG GCA CCC CTG GCT ATC AGG CCA GCG TCG ACT  
pro thr pro thr thr ala leu ala arg trp ala pro leu ala ile arg pro ala ser thr  
123/41  
ATG TGG TAA ACA CAC TGC GCA ACA GCG GTT TTG ATG TGC AAA CCC CGG AGT TCT CCG CTC  
met trp OCH thr his cys ala thr ala val leu met cys lys pro arg ser ser pro leu  
183/61  
GCG TGT TCA AGG CCG AAA AAG GGG TGG TGA CCC TCG GCG GCA ACA CCG TGG AGG CGA GGG  
ala cys ser arg pro lys lys gly trp OPA pro ser ala ala thr pro trp arg arg gly  
243/81  
CGC TCG AGT ACA GCG TCG GCA CAC CGC CGG ACG GGG TGA CCG GCC CGC TGG TGG CTC CCC  
arg ser ser thr ala ser ala his arg arg thr gly OPA arg ala arg trp trp leu pro  
303/101  
CCG CCG ACG ACA GTC CGG GCT GCA GTC CGT CGG ACT ACG ACA GGC TGC CCG TGT CCG GTG  
pro pro thr thr val arg ala ala val arg arg thr thr thr gly cys arg cys pro val  
363/121  
CGG TGG TGC TGG TAG ATC  
arg trp cys trp AMB ile

SEQ ID N° 19C

FIGURE 19C

partie de la séquence nucléotidique de seq19A

1/1  
31/11  
CTA TCG AAA CTC CAG GAG ATC GCG AAA GCG AAG GAG GTC A T TCG GCG ATG GCG AAT CCG  
leu ser lys leu gln asp ile ala asp ala asp arg gly thr ala ala val gly thr pro  
61/21  
GGC TAT CAG GCC AGC CTC GAC TAT GTG GTA AAG ACA CCG GCG AAT AGC GGT CTT GAT GTC  
gly tyr gln ala ser val asp tyr val val arg thr leu thr ala ser gly phe arg val  
121/41  
CAA AGC CAG GAG TTC TTA GTC GCG ATG TT AAG GCG TAA AAA GCG GTC CCG AGC CCG GCG  
gln thr pro gln phe ser ala arg val phe tyr ala ala lys gly val val thr leu gly  
181/61  
GCG AAG AAT TCG TAA GCG AAG TCG CTC GAT TAT AAT GCG TAA AAA GCG GTC CCG AGC CCG GCG  
gly asp thr val ala ala thr ala asp thr leu thr ala ala lys gly val val thr leu gly  
241/81  
ACG GGT CAG CTC GTC GTC GTT TCG GCG GCG AAG GAG AAT GCG GCG TAA AAA GCG GTC CCG AGC CCG GCG  
thr gly pro leu val ala ala pro ala arg arg thr leu thr ala ala lys gly val val thr leu gly  
301/101  
GCG GGT CAG CTC GTC GTC GTT TCG GCG GCG AAG GAG AAT GCG GCG TAA AAA GCG GTC CCG AGC CCG GCG  
thr gly pro leu val ala ala pro ala arg arg thr leu thr ala ala lys gly val val thr leu gly  
361/121  
GCG GGT CAG CTC GTC GTC GTT TCG GCG GCG AAG GAG AAT GCG GCG TAA AAA GCG GTC CCG AGC CCG GCG  
thr gly pro leu val ala ala pro ala arg arg thr leu thr ala ala lys gly val val thr leu gly

69/185

[illegible]

SEQ ID N° 19B'

FIGURE 19B'

[illegible]

70/185

sequence Rv0418 prédite par Cole et al. (Nature 393:537-544) et contenant seq19A'

```

1/1                               31/11
atg gtg aac aaa tcc agg atg atg ccg gcc gtg ctg gcc gtg gct gtg gtc gtc gca ttc
Met val asn lys ser arg met met pro ala val leu ala val ala val val ala phe
61/21                               91/31
ctg acg acg gcc tgt atc cgg tgg tct acg cag tcg cgg ccc gtt gtt aac gcc ccc gct
leu thr thr gly cys ile arg trp ser thr gln ser arg pro val val asn gly pro ala
121/41                               151/51
gcc gca gag ttc gcc gtt gcc ttg cgc aac cgg gtg agc acc gac gcc atg atg gcc cac
ala ala glu phe ala val ala leu arg asn arg val ser thr asp ala met met ala his
181/61                               211/71
cta tcg aaa ctg cag gac atc gcc aac gcc aac gac gcc act cgc gcc gtg gcc acc cct
leu ser lys leu gln asp ile ala asn ala asn asp gly thr arg ala val gly thr pro
241/81                               271/91
ggc tat cag gcc agc gtc gac tat gtg gta aac aca ctg cgc aac agc ggt ttt gat gtg
gly tyr gln ala ser val asp tyr val val asn thr leu arg asn ser gly phe asp val
301/101                               331/111
aaa acc ccg gag ttc tcc gct cgc gtg ttc aag gcc gaa aaa ggg gtg gtg acc ctc gcc
gln thr pro glu phe ser ala arg val phe lys ala glu lys gly val val thr leu gly
361/121                               391/131
ggc aac acc gtg gag gcc agg gcc ctc gag tac agc ctc gcc aca ccg ccg gac ggg gtg
gly asn thr val glu ala arg ala leu glu tyr ser leu gly thr pro pro asp gly val
421/141                               451/151
acg gcc ccg ctg gtg gct gcc ccc gcc gac gac agt ccg gcc tgc agt ccg tcg gac tac
thr gly pro leu val ala ala pro ala asp asp ser pro gly cys ser pro ser asp tyr
481/161                               511/171
gac agg ctg ccg gtg tcc ggt gcc gtg gtg ctg gta gat cgc gcc gtc tgt cct ttt gcc
asp arg leu pro val ser gly ala val val leu val asp arg gly val cys pro phe ala
541/181                               571/191
cag aag gaa gac gca gcc gcc cag cgg ggt gcc gtg gcc ctg atc att gct gac aac atc
gln lys glu asp ala ala ala gln arg gly ala val ala leu ile ile ala asp asn ile
601/201                               631/211
gac gag cag gcc atg gcc gcc acc ctg gcc gct aat acc gac gtc aag atc ccc gtg gtg
asp glu gln ala met gly gly thr leu gly ala asn thr asp val lys ile pro val val
661/221                               691/231
agt gtc acc aag tcg gtc gga ttc cag cta cgc gga cag tct ggg cca acc acc gtc aag
ser val thr lys ser val gly phe gln leu arg gly ala ser gly pro thr thr val lys
721/241                               751/251
gac aag gcc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc
leu thr ala ser thr ala ser phe lys ala ala ala val ile ala thr thr lys thr gly
781/261                               811/271
gac atc gcc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc
ser ser ala asn val val met ala gly ala ala leu asp ser val pro ala gly gly gly
841/281                               871/291
acc aac gcc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc
ile asn asp asn gly ser gly val ala ala val leu ala thr ala val ala leu gly ala
901/301                               931/311
gac atc gcc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc
thr thr ala thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

```

FIN DE LA PAGE

TABLEAU DE REMPLACEMENT (REGLE 26)

71/185

1021/341	1051/351
tat ctg aac ttc gac atg ttg gcg tcg ccg aac ccg ggt tac ttc acc tac gac ggt gac	
tyr leu asn phe asp met leu ala ser pro asn pro gly tyr phe thr tyr asp gly asp	
1081/361	1111/371
cag tcg ctg ccg cta gac gcc cgc ggt cag ccg gtg gtg ccc gaa gcc tcg gcc ggt atc	
gln ser leu pro leu asp ala arg gly gln pro val val pro glu gly ser ala gly ile	
1141/381	1171/391
gag cgc acg ttc gtc gcc tat ctg aag atg gcc gcc aag acc gcg cag gac acc tcg ttc	
glu arg thr phe val ala tyr leu lys met ala gly lys thr ala gln asp thr ser phe	
1201/401	1231/411
gac ggt cgg tcc gac tac gac gcc ttc acg ctg gcg ggt atc cct tcg ggt gcc ctg ttc	
asp gly arg ser asp tyr asp gly phe thr leu ala gly ile pro ser gly gly leu phe	
1261/421	1291/431
tcc gcc gct gag gtc aag aag tcc gcc gag caa gcc gag ctg tcg gcc gcc acc gcc gac	
ser gly ala glu val lys lys ser ala glu gln ala glu leu trp gly gly thr ala asp	
1321/441	1351/451
gag cct ttc gat ccc aac tat cac cag aag acc tac acc ctg gac cat atc gac cgc acc	
gln pro phe asp pro asn tyr his gln lys thr asp thr leu asp his ile asp arg thr	
1381/461	1411/471
gcg ctg ggt atc aac gcc gct gcc gtc gcg tac ccg gtg ggt ttg tat gcg cag gac ctg	
ala leu gly ile asn gly ala gly val ala tyr ala val gly leu tyr ala gln asp leu	
1441/481	1471/491
ggc gcc ccc aac ggg gtt ccg gtc atg gcg gac cgc acc cgc cac ctg att gcc aaa ccg	
gly gly pro asn gly val pro val met ala asp arg thr arg his leu ile ala lys pro	
1501/501	
tga	
OPA	

SEQ ID N° 190 (suite)

FIGURE 190 (suite)

PLATE 10. 6179A. FLEMING 200



73/185

1021/341 1051/351  
 tgg ggc gcc gag gaa ttc ggc ctg att ggg tca cga aac tac gtc gag tcg ctg gac atc  
 trp gly ala glu glu phe gly leu ile gly ser arg asn tyr val glu ser leu asp ile  
 1081/361 1111/371  
 gac gcg ctc aaa ggc atc gcg ctg tat ctg aac ttc gac atg ttg gcg tcg ccg aac ccg  
 asp ala leu lys gly ile ala leu tyr leu asn phe asp met leu ala ser pro asn pro  
 1141/381 1171/391  
 ggt tac ttc acc tac gac ggt gac cag tcg ctg ccg cta gac gcc cgc ggt cag ccg gtg  
 gly tyr phe thr tyr asp gly asp gln ser leu pro leu asp ala arg gly gln pro val  
 1201/401 1231/411  
 gtg ccc gaa ggc tcg gcc ggt atc gag cgc acg ttc gtc gcc tat ctg aag atg gcc gcc  
 val pro glu gly ser ala gly ile glu arg thr phe val ala tyr leu lys met ala gly  
 1261/421 1291/431  
 aag acc gcg cag gac acc tcg ttc gac ggt cgg tcc gac tac gac ggc ttc acg ctg gcg  
 lys thr ala gln asp thr ser phe asp gly arg ser asp tyr asp gly phe thr leu ala  
 1321/441 1351/451  
 ggt atc cct tcg ggt ggc ctg ttc tcc ggc gct gag gtc aag aag tcc gcc gag caa gcc  
 gly ile pro ser gly gly leu phe ser gly ala glu val lys lys ser ala glu gln ala  
 1381/461 1411/471  
 gag ctc tgg ggc gcc acc gcc gac gag cct ttc gat ccc aac tat cac cag aag aca gac  
 glu leu trp gly gly thr ala asp glu pro phe asp pro asn tyr his gln lys thr asp  
 1441/481 1471/491  
 acc ctg gac cat atc gac cgc acc gcg ctc ggt atc aac gcc gct gcc gtc gcg tac gcg  
 thr leu asp his ile asp arg thr ala leu gly ile asn gly ala gly val ala tyr ala  
 1501/501 1531/511  
 gtg ggt ttg tat gcg cag gac ctc gcc gcc ccc aac ggg gtt ccg gtc atg gcg gac cgc  
 val gly leu tyr ala gln asp leu gly gly pro asn gly val pro val met ala asp arg  
 1561/521  
 acc cgc cac ctg att gcc aaa ccg tga  
 thr arg his leu ile ala lys pro opa

SEQ ID N° 19F (suite)

FIGURE 19F (suite)

31/11  
 GAA GAA AGT GGT GCG GGA GAA TTG AGT TCG GCG GAA AAA GAA GAA GAA GAA GAA GAA GAA GAA  
 Arg Asp Ser Gly Ala Gly His Leu Ser Ser Ala Ala Lys Arg Ala Arg Val Ala Arg Phe  
 61/21 91/31  
 GAG GGT GCG GAA GCG ACC GCG GGT GAG GGT GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA  
 His Gly Val Gly Leu Thr Phe Gly Ala Gly Thr Ala Lys Leu Arg Ala Arg Val  
 121/41 151/51  
 GAG GAG GCG GAA AAA GCG GGT TCG GAA GGT GGT GAA GAA GAA GAA GAA GAA GAA GAA GAA  
 Gly Arg Leu Ala Thr Val Gly Phe Asp Val Gly Asp Asp Pro Ser Phe Thr Lys Arg Ser  
 181/61 211/71  
 TGA GGT GCG GGT TTT GGT GAA TCG GGA GCG ACC GCG GGT AAT GAA GCG GGT AAA GGT GCG  
 Stop Gly Arg Ala Lys Gly Tyr His Gly Thr Ala Gly Thr Ala Gly Thr Ala Gly Thr Ala  
 241/91 271/101  
 GAA GAA GCG GGT GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA  
 Glu Ala Glu Asp Thr Ala Asp GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA  
 301/111 331/121

74/185

32/11  
 GAG ACA GTG GTG CGG GAC ACT TGA GTT CGG CTG CTA ACG ACG CCA GAG TCG CCC GCT TCC  
 glu thr val val arg asp thr OPA val arg leu leu thr thr pro glu ser pro ala ser  
 62/21  
 GCG GTG TGG GAC TCA CGT TCG GTG AGG GTA CAG CGG ACC TTC GAG CAC GCA ATA TCG TGG  
 ala val trp asp ser arg ser val arg val gln arg thr phe glu his ala ile ser trp  
 122/41  
 GCC GGC TGG CAA CCG TCG GTT TCG ACG TTG GTG ACG ACC CCT CGT TCA TGA ATC GTT CTT  
 ala gly trp gln pro ser val ser thr leu val thr thr pro arg ser OPA ile val leu  
 182/61  
 GAG CTC CCC GTT TTG CTG GAT GCC CAG GCA CCG CCG GTA CTG CTG CCG TTA AGC TTG TCG  
 glu leu pro val leu leu asp ala gln ala pro pro val leu leu arg leu ser leu ser  
 242/81  
 CAC ATG GTG CCG GCA GGG AGG AAC AGT GGG CAA GCA GCT AGC CCG GCT CCG CCG GCT GGT  
 his met val pro ala gly arg asn ser gly gln ala ala ser arg ala arg arg ala gly  
 302/101  
 CGG TGC GTG CAT GCT CSC AGC CCG ATG CAG CAA CGT GGT CGA CCG GAC CCG CGT GCG TGC  
 arg cys val his ala arg ser arg met his gln arg gly arg arg asp arg arg gly cys  
 362/121  
 CGA CAA ATC CCG ACC ACT GCA TCA GGA TC  
 arg gln ile arg thr thr ala ser gly

SEQ ID N° 20B

FIGURE 20B

33/11  
 AGA CAG TGG TGC GGG ACA CTT GAG TTC GGC TCG TAA CGA CCG CAG AGT CCG CCG GTT CCG  
 arg gln trp cys gly thr leu gln phe gly cys CCH arg arg gln ser arg pro leu pro  
 63/21  
 CCG TGT GCG ACT CAG GTT CCG TGA GCG TAT ATT GGA ATT TTT AGT ATG CAA TAT CCG GAT  
 ala cys gly thr his val arg CCA gly tyr ser gly pro ser ser thr ala tyr arg gly  
 123/41  
 CCG GGT CCG AAT CCG CCG TTT CCA CCG TCG TTA CCA ATT CCG GTT CAT GAA TAT TTT TTT  
 pro ala gly asp arg ala pro ala arg trp CCA arg pro leu val his ala ser phe leu  
 183/61  
 AGC TCG CCG TTT TCG TGG ATG CCG AGG CAG CCG CCG TAT TTT TTT CTT TAA GGT TGT CCG  
 ser ser his phe cys thr pro pro arg his arg ala tyr cys cys ala CCH ala cys arg  
 243/81  
 ATA TCG TAT CCG CAG GGA GGA AAT GTC GGT AAG TAT CTA GGT TTT TTG GGT TTT TTT TTT  
 thr thr thr arg gln gly thr thr val gly lys ala pro ala ala leu ala ala ala ala  
 303/101  
 ATT CCG TAT ATT CTT TTA ATT TTA TTA ATT TTA TTA ATT TTA TTA ATT TTA TTA ATT TTA  
 arg cys thr att ctt tta att tta tta att tta tta att tta tta att tta tta att tta  
 363/121

SEQ ID N° 21B

SEQUENCE ALIGNMENT OF FIGURE 20B

75/185

partie de la séquence nucléotidique de seq20A

```

1/1                               31/11
TGT GGG ACT CAC GTT CGG TGA GGG TAC AGC GGA CCT TCG AGC ACG CAA TAT CGT GGG CCG
cys gly thr his val arg OPA gly tyr ser gly pro ser ser thr gln tyr arg gly pro
61/21                               91/31
GCT GGC AAC CGT CGG TTT CGA CGT TGG TGA CGA CCC CTC GTT CAT GAA TCG TTC TTG AGC
ala gly asn arg arg phe arg arg trp OPA arg pro leu val his glu ser phe leu ser
121/41                               151/51
TCC CCG TTT TGC TGG ATG CCC AGG CAC CGC CGG TAC TGC TGC GGT TAA GCT TGT CGC ACA
ser pro phe cys trp met pro arg his arg arg tyr cys cys ala OCH ala cys arg thr
181/61                               211/71
TGG TGC CGG CAG GGA GGA ACA GTG GGC AAG CAG CTA GCC GCG CTC GCC GCG CTG GTC GGT
trp cys arg gln gly gly thr val gly lys gln leu ala ala leu ala ala leu val gly
241/81                               271/91
GGC TGC ATG CTC GCA GGC GGA TGC ACC AAC GTG CTC GAG GGG ACC GCG GTG GCT GCG GAC
ala cys met leu ala ala gly cys thr asn val val asp gly thr ala val ala ala asp
301/101
AAA TCC GGA CCA CTG CAT CAG GAT C
lys ser gly pro leu his gln asp

```

SEQ ID N° 20A'

FIGURE 20A'

```

1/1                               31/11
GTG GGA CTC ACG TTC GGT GAG GGT ACA GCG GAG GTT CGA GGA GGC AAT ATC GTG GGC CGG
val gly leu thr phe gly glu gly thr ala asp leu arg ala arg asn ile val gly arg
61/21                               91/21
CTG GCA ACC CTC GGT TTC GAG GTT GGT GAG GAG GGT TAT TTT ATG AAT GGT TCT TGA GGT
leu ala thr val gly phe asp val gly asp arg pro arg phe asp asp arg ser OPA ala
121/41                               151/51
GGC GGT TTT GGT GGA TGC GCA GGC ACC GCG GAG AAT GGT GAG GTT AAG GGT GTC GTA CAT
pro arg phe ala gly cys pro gly thr ala gly thr ala ala leu lys leu val ala his
181/61                               211/71
GGT GCG GGC AAG GAG GAA GAG TGC GCA AAT AAT GAG GAG GAG GAG GGT TGT TAT GTT
gly ala gly arg glu glu glu trp ala ser ser AME pro arg pro arg trp ser val
241/81                               271/91
GCT GCA TAT TAT GAG GCG GAG GAA GAA GAA AAT TAT TAT GAG GAG GAG GAG GGT TGT TAT
thr ala thr thr thr pro asp ala pro thr thr thr thr thr thr thr thr thr thr
301/101
AAT GCG GAG AAT TAT AAT AAG AAT
asp pro asp ala cys ala arg ala

```

SEQ ID N° 20B'

76/185

```

1/1                               31/11
GTG TGG GAC TCA CGT TCG GTG AGG GTA CAG CGG ACC TTC GAG CAC GCA ATA TCG TGG GCC
val trp asp ser arg ser val arg val gln arg thr phe glu his ala ile ser trp ala
61/21                               91/31
GGC TGG CAA CCG TCG GTT TCG ACG TTG GTG ACG ACC CCT CGT TCA TGA ATC GTT CTT GAG
gly trp gln pro ser val ser thr leu val thr thr pro arg ser opa ile val leu glu
121/41                               151/51
CTC CCC GTT TTG CTG GAT GCC CAG GCA CCG CCG GTA CTG CTG CGC TTA AGC TTG TCG CAC
leu pro val leu leu asp ala gln ala pro pro val leu leu arg leu ser leu ser his
181/61                               211/71
ATG GTG CCG GCA GGG AGG AAC AGT GGG CAA GCA GCT AGC CGC GGT CGC CGC GGT GGT CGG
met val pro ala gly arg asn ser gly gln ala ala ser arg ala arg arg ala gly arg
241/81                               271/91
TGC GTG CAT GCT CGC AGC CGG ATG CAC CAA CGT GGT CGA CGG GAC CGC CGT GGC TGC CGA
cys val his ala arg ser arg met his gln arg gly arg arg asp arg arg gly cys arg
301/101
CAA ATC CCG ACC ACT GCA TCA GGA TC
gln ile arg thr thr ala ser gly

```

SEQ ID N° 20C'

FIGURE 20C'

```

séquence RV3576 prédite par Cole et al. (Nature 393:537-544) et contenant seq20A'
1/1                               31/11
atg ggc aag cag cta gcc gcg ctc gcc gcg ctg gtc ggt gcg tgc atg ctc gca gcc gga
met gly lys gln leu ala ala leu ala ala leu val gly ala cys met leu ala ala gly
61/21                               91/31
tgc acc aac gtg gtc gac ggg acc gcc gtg gct gcc gac aaa tcc gga cca ctg cat cag
cys thr asn val val asp gly thr ala val ala ala asp lys ser gly pro leu his gln
121/41                               151/51
gat ccg ata ccg gtt tca gcg ctt gaa ggg ctg ctt ctc gac ttg acc cag atc aat gcc
asp pro ile pro val ser ala leu glu gly leu leu leu asp leu ser gln ile asn ala
181/61                               211/71
ggc ctg ggt gcg ata tgc atg aag gtg tgg ttc aac gtc aag gaa ata tgg gac tgg acc
ala leu gly ala thr ser met lys val trp phe asn ala lys ala met trp asp trp ser
241/81                               271/91
aag acc gtg gcc gac aag aat tcc ctg cct atc gac gac cca gaa cag gaa aag gtc tat
lys ser val ala asp lys asn cys leu ala ile asp gly pro ala gln glu lys val tyr
301/101                               331/111
gcc tgc acc ggg tgg acc gat ata ggc ggc gaa ggc gtc gtc gtc gtc gtc gtc gtc gtc
ala gly thr gty trp thr ala met gln gty ala gln ala gln ala gln ala gln ala gln
361/121                               391/131
gac aac gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc
lys lys arg asp his trp ala ile gln ala tal val gty gln gln gln gln gln gln gln
421/141                               451/151
gac ggc ttc ttc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc
glu glu phe tyr ser ser ser val gln ser trp ser ser ggc ser asn ala arg phe val
481/161                               511/171
gaa ggc acc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc
ala val thr pro gly gln asp asp ala ala ttc thr val ala val val val val val val
541/181                               571/191
gac gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc
gty met leu ser ser ser ser val gln gln gln gln gln gln gln gln gln gln gln gln
601/201

```

FIG. 20C' (continued)

SEQUENCE 20C' COMPLEMENT (REV. 1.0)

77/185

ORF d'après Cole et al. (Nature 393:537-544) et contenant Rv3576

```

1/1                               31/11
taa gct tgt cgc aca tgg tgc cgg cag gga gga aca qtg ggc aag cag cta gcc gcg ctc
OCH ala cys arg thr trp cys arg gln gly gly thr val gly lys gln leu ala ala leu
61/21                               91/31
gcc gag ctg gtc ggt gcg tgc atg ctc gca gcc gga tgc acc aac gtg gtc gac ggg acc
ala ala leu val gly ala cys met leu ala ala gly cys thr asn val val asp gly thr
121/41                               151/51
gcc gtg gct gcc gac aaa tcc gga cca ctg cat cag gat ccg ata ccg gtt tca gcg ctt
ala val ala ala asp lys ser gly pro leu his gln asp pro ile pro val ser ala leu
181/61                               211/71
gaa ggg ctg ctt ctc gac ttg agc cag atc aat gcc gcc ctg gat gcg aca tgg atg aag
glu gly leu leu leu asp leu ser gln ile asn ala ala leu gly ala thr ser met lys
241/81                               271/91
gtg tgg ttc aac gcc aag gca atg tgg gac tgg agc aag agc gtg gcc gac aag aat tgc
val trp phe asn ala lys ala met trp asp trp ser lys ser val ala asp lys asn cys
301/101                               331/111
ctg gct atc gac ggt cca gca cag gaa aag gtc tat gcc gcc acc ggg tgg acc gat atg
leu ala ile asp gly pro ala gln glu lys val tyr ala gly thr gly trp thr ala met
361/121                               391/131
cgc gcc caa cgg ctg gat gac agc atc gat gac tcc aag aaa cgc gac cac tac gcc att
arg gly gln arg leu asp asp ser ile asp asp ser lys lys arg asp his tyr ala ile
421/141                               451/151
caa gcg gtc gtc gcc ttc ccg acc gca cat gat gcc gag gag ttc tac agc tcc tgg gtg
glu ala val val gly phe pro thr ala his asp ala glu glu phe tyr ser ser ser val
481/161                               511/171
caa agc tgg agc agc tgc tgc aac cgc cgg ttt gtc gaa gtc acc ccc gga cag gac gac
glu ser trp ser ser cys ser asn arg arg phe val glu val thr pro gly gln asp asp
541/181                               571/191
gcc gcc tgg act gtg gct gac gtt gtc aac gac aac ggc atg atc agt agc tgg cag gtt
ala ala trp thr val ala asp val val asn asp asn gly met leu ser ser ser ala val
601/201                               631/211
cag gaa ggc ggc gac gaa tgg acc tgc cag ctt gtc tgc act ggc cgc aac aac gtc aat
glu glu gly gly asp gly trp thr cys gln arg ala leu thr ala arg asn asn val thr
661/221                               691/231
atc gac att gtc arg ttc ser tat agc caa ttc atc ttc ggc ggc gtc gtc gtc gtc
ala asp ala val thr cys ala tyr ser ala thr agc leu val ala ala gly ala ala ala
721/241
gac ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala

```

SEQ ID N° 20F

FIGURE 20F

78/185

1/1 31/11  
 GTC CTG GTC GCC GCG CAA CTG GCC GGT CCC GAT GGA AAG TGT TCA CGA TCG CGC TTC TGC  
 val leu val ala ala gln leu ala gly pro asp gly lys cys ser arg ser arg phe cys  
 61/21 91/31  
 CGC TGG TAG TGG CGA TGG TGT TAG CAG GAT TGC GGG TCG AGG CTG CGA TGG CCA GCA CCA  
 arg trp AMB trp arg trp cys AMB gln asp cys gly ser arg leu arg trp pro ala pro  
 121/41 151/51  
 GCG GCC TGC GGC TGG TCG CCG GCG CCG AAA TGA TAC CCG CGA TCA CGA AAT ACA TGT  
 ala ala cys gly trp ser pro arg ala pro lys OPA tyr pro arg ser arg asn thr cys  
 181/61 211/71  
 CGG CGC TGG ACG TCG CCG TGC TGG CCA GCT CGA CCG GAC ACG ATG TGG AGG GGG CGC AGA  
 arg arg trp thr ser pro cys trp pro ala arg pro asp thr met tip arg gly arg arg  
 241/81 271/91  
 AAA ACT TCA CCG CCC GCA AGT ACG AGC TGC AGA CCG GAC TGG CCG ACA CCG ACG TCA TCG  
 lys thr ser pro pro ala ser thr ser cys arg arg asp trp pro thr pro thr ser ser  
 301/101 331/111  
 CAG ACG TGC GGT CCG GAG TGA ACA CSC TGC TCA ACG GCG GTC AGG CGC TGC TGG ATA AGA  
 gln thr cys gly arg glu OPA thr arg cys ser thr ala val arg arg cys trp ile arg  
 361/121  
 TGC TGG CCG ACA GCA TCG GCT TGC GGG ATC  
 cys trp pro thr ala ser ala cys gly ile

SEQ ID N° 21A

FIGURE 21A

32/11  
 TCC TGS TCG CCG CCG AAC TCG CCG GTC CCG ATG GAA AGT GTT CAC GAT CGC GCT TCT GCG  
 ser trp ser pro arg asn trp pro val pro met glu ser val his asp arg ala ser ala  
 62/21 92/31  
 GGT GGT AGT GCG GAT GGT GTT AGC AGG ATT GCG GGT CGA GCG TGC GAT GCG CAG CAC CAG  
 ala gly ser gly asp gly val ser arg ile ala gly arg gly cys asp gly gln his gln  
 122/41 152/51  
 CGG GGT GCG GGT GGT GCG CCG CCG GCG CCG GCA AAT GAT ACG GTC GAT CAC GAA ATA CAT GTT  
 arg pro ala ala gly arg arg ala arg ala asp asp thr arg asp his gln ile his val  
 182/61 212/71  
 GCG GGT CGA GGT GGT GGT GGT GGT GCG CAG CTC GAG CCG AAA CCA TAT TGA GCG GGT GCA GAA  
 try ala gly arg his arg ala gly gln leu arg arg thr arg cys gly gly gly ala ala  
 242/81 272/91  
 AAA GTT GAT GAT GTT GAA TTA CGA GGT GGA GAT GCG AAT GAT TCA TAT CGA GGT GAT GAT  
 lys leu his arg pro ala val arg ala ala asp ala thr gly ala his arg arg his arg  
 302/101 332/111  
 AAA GGT GAT GTT GGT GAT GAA GAT GGT GAT GAA GCG CCG TCA TAT GTT GTT GCA TAT GAT  
 arg arg ala val arg arg ala his ala ala ala arg his arg arg arg ala gly thr arg  
 362/121

79/185

33/11  
 CCT GGT CGC CGC GCA ACT GGC CGG TCC CGA TGG AAA GTG TTC ACG ATC GCG CTT CTG CCG  
 pro gly arg arg ala thr gly arg ser arg trp lys val phe thr ile ala leu leu pro  
 63/21  
 CTG GTA GTG GCG ATG GTG TTA GCA GGA TTG CGG GTC GAG GCT GCG ATG GCC AGC ACC AGC  
 leu val val ala met val leu ala gly leu arg val glu ala ala met ala ser thr ser  
 123/41  
 GGC CTG CGG CTG GTC GCC GCG CGC GCC GAA ATG ATA CCC GCG ATC ACG AAA TAC ATG TCG  
 gly leu arg leu val ala ala arg ala glu met ile pro ala ile thr lys tyr met ser  
 183/61  
 GCG CTG GAC GTC GCC GTG CTG GCC AGC TCG ACC GGA CAC GAT GTG GAG GGG GCG CAG AAA  
 ala leu asp val ala val leu ala ser ser thr gly his asp val glu gly ala gln lys  
 243/81  
 AAC TTC ACC GCC CGC AAG TAC GAG CTG CAG ACC CGA CTG GCG GAC ACC GAC GTC ATC CCA  
 asn phe thr ala arg lys tyr glu leu gln thr arg leu ala asp thr asp val ile ala  
 303/101  
 GAC GTG CGG TCG GGA GTG AAC ACG CTG CTC AAC GGC GGT CAG GCG CTG CTG GAT AAG ATG  
 asp val arg ser gly val asn thr leu leu asn gly gly gln ala leu leu asp lys met  
 363/121  
 CTG GCC GAC AGC ATC GGC TTG CCG GAT C  
 leu ala asp ser ile gly leu arg asp

SEQ ID N° 21C

FIGURE 21C

partie de la séquence nucléotidique de seq1A

1/1  
 ACG ATC GCG CTT CTG CCG CTG GTA GTG GCG ATG CTG TTA GCA GGA TTG CCG GTC GAG GCG  
 thr ile ala leu leu pro leu val val ala met val leu ala gly leu arg val gln ala  
 61/21  
 GCG ATG GCG AGC ACC ACC GGC CTG CCG CTG CTG CTG CTG CTG GAA ATG ATA GCG GCG  
 ala met ala ser thr ser gly leu arg leu val ala ala thr ala gln met ala val ala  
 121/41  
 ATC ACC AAA TAC ATG TTA GCG CTG GAG CTG ACC CTG CTG CTG CTG CTG GAA TAC GAG  
 ile thr lys tyr met thr ala leu arg val ala val ala ala thr thr thr gly thr arg  
 181/61  
 CTG GAG GCG GTC ACG AAA AAC TTA A GCG TAC GAG GTC TAC ACC GCG CTG GTC  
 val glu gly ala ala lys asn phe thr ala arg lys tyr thr leu gln thr thr thr ala  
 241/81  
 GAG ACT GAG CTG ATC GGA AAC GTG GCG TCG GAA GCG AAC ACC GTC AAC GAG GCG GAG  
 asp thr asp val ile ala arg val thr thr thr val thr thr thr thr thr thr thr thr  
 301/101  
 GAG TTA GCG GTC AAC ACC GCG GTC GAG GCG GAG GCG GAG GCG GAG GCG GAG GCG GAG GCG  
 asp thr asp val ile ala arg val thr thr thr thr thr thr thr thr thr thr thr thr thr

80/185

```

1/1                               31/11
CGA TCG CGC TTC TGC CGC TGG TAG TGG CGA TGG TGT TAG CAG GAT TGC GGG TCG AGG CTG
arg ser arg phe cys arg trp AMB trp arg trp cys AMB gln asp cys gly ser arg leu
61/21                               91/31
CGA TGG CCA GCA CCA GCG GCC TGC GGC TGG TCG CCG CGC GCG CCG AAA TGA TAC CCG CGA
arg trp pro ala pro ala ala cys gly trp ser pro arg ala pro lys OPA tyr pro arg
121/41                               151/51
TCA CGA AAT ACA TGT CGG CGC TGG ACG TCG CCG TGC TGG CCA GCT CGA CCG GAC ACG ATG
ser arg asn thr cys arg arg trp thr ser pro cys trp pro ala arg pro asp thr met
181/61                               211/71
TGG AGG GGG CGC AGA AAA ACT TCA CCG CCG GCA AGT ACG AGC TGC AGA CGC GAC TGG CCG
trp arg gly arg arg lys thr ser pro pro ala ser thr ser cys arg arg asp trp pro
241/81                               271/91
ACA CCG ACG TCA TCG CAG ACG TGC GGT CCG GAG TGA ACA CCG TGC TCA ACG GCG GTC AGG
thr pro thr ser ser gln thr cys gly arg glu OPA thr arg cys ser thr ala val arg
301/101                               331/111
CGC TGC TGG ATA AGA TGC TGG CCG ACA GCA TCG GCT TGC GGG ATC
arg cys trp ile arg cys trp pro thr ala ser ala cys gly ile

```

SEQ ID N° 21B'

FIGURE 21B'

```

1/1                               31/11
CAC GAT CGC GCT TGT GCG GGT AGT GCG CAT GGT GTT AGC AGC ATT GCG GGT CGA GCG
his asp arg ala ser ala ala gly ser gly asp gly val ser arg ile ala gly arg gly
61/21                               91/31
TGC GAT GGC CAG CAC CAG CCG CCT GCG GGT GGT GGT GGT GCG GGT CGA AAT GAT ACG CGC
cys asp gly gln his gln arg pro ala ala gly arg arg ala arg arg arg asp thr arg
121/41                               151/51
GAT CAG GAA ATA GAT GTT GCG GCT GGA GAT GAT GGT GGT GAT GAT GAT GAT GAT GAT GAT
asp his ala ile his val gly ala gly arg arg arg ala gly thr leu arg arg thr arg
181/61                               211/71
TGT GGA GCG GCG GCA GAA AAA GTT CAT GAT GAT GAT GAA GAA GAA GAT GAA GAT GAT GAT
lys gly gly gly ala gln lys leu his arg arg arg val arg ala ala arg ala thr gly
241/81                               271/91
GAA CAG GAA GGT GAT GAT GAA GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
arg his arg arg his arg arg arg val gly arg arg arg ala ala gln arg arg arg
301/101                               331/111
GAT GGT GAT GGA TAA GAT GAT GGT GGT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
gly ala ala gly GCH arg ala gly arg ala his arg arg arg arg

```



81/185

séquence Rv3365c prédite par Cole et al. (Nature 393:537-544) et contenant Seq21A'

```

1/1                               31/11
gtg acc atg ttc gcc cgc ccg acc atc ccg gtc gcg gcg gcc gct tct gat att tcc gcc
val thr met phe ala arg pro thr ile pro val ala ala ala ala ser asp ile ser ala
61/21                               91/31
ccg gct caa ccg gcc cgc gcc aaa cct cag caa cgc ccg ccg tcc tgg tgg ccg cgc aac
pro ala gln pro ala arg gly lys pro gln gln arg pro pro ser trp ser pro arg asn
121/41                               151/51
tgg ccg gtc cga tgg aaa gtg ttc acg atc gcg ctt ctg ccg ctg gta gtg gcg atg gtg
trp pro val arg trp lys val phe thr ile ala leu leu pro leu val val ala met val
181/61                               211/71
tta gca gga ttg ccg gtc gag gct gcg atg gcc agc acc agc gcc ctg ccg ctg gtc gcc
leu ala gly leu arg val glu ala ala met ala ser thr ser gly leu arg leu val ala
241/81                               271/91
gcg cgc gcc gaa atg ata ccc gcg atc acg aaa tac atg tgg gcg ctg gac gtc gcc gtg
ala arg ala glu met ile pro ala ile thr lys tyr met ser ala leu asp val ala val
301/101                               331/111
ctg gcc agc tgg acc gga cac gat gtg gag ggg gcg cag aaa aac ttc acc gcc cgc aag
leu ala ser ser thr gly his asp val glu gly ala gln lys asn phe thr ala arg lys
361/121                               391/131
tac gag ctg cag acg cga ctg gcc gac acc gac gtc atc gca gac gtg ccg tgg gga gtg
tyr glu leu gln thr arg leu ala asp thr asp val ile ala asp val arg ser gly val
421/141                               451/151
aac acg ctg ctg aac gcc ggt cag gcg ctg ctg gat aag gtg ctg gcc gac agc atc gcc
asn thr leu leu asn gly gly gln ala leu leu asp lys val leu ala asp ser ile gly
481/161                               511/171
ttg ccg gat ccg gtc acc gcc tac gcg ccg ctg ctg ttg acg gcc cag aac gtg att gac
leu arg asp arg val thr ala tyr ala pro leu leu leu thr ala gln asn val ile asp
541/181                               571/191
gcg tgg gtg ccg gtt gac agc gag caa atc cga acc cag gtg cag ggt ttg agc cga gcc
ala ser val arg val asp ser glu gln ile arg thr gln val gln gly leu ser arg ala
601/201                               631/211
gtt gcc gcc cgc gcc cag atg acg atg cag gat atc ctg gtc act gln gln gln gac att
val gly ala arg gly gln met thr met gln gln ile leu val thr arg gly ala asp leu
661/221                               691/231
gcc gaa ccg cag atg acc agc agc atg gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc
ala gln pro ala leu arg ser ala met val thr leu ala gly thr gln pro ser thr leu
721/241                               751/251
gtt gcc atg acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc
pro gly met ser ala ala leu gly ala gly ser pro arg thr lys ala leu gln gln ala
781/261                               811/271
gtt gtc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc
met val thr arg met ala ile met ser asp pro ala val ala leu val ala ala pro gln
841/281                               871/291
gtt gcc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc
leu leu his ser ile ala ala thr arg arg ala leu ala ala ala ala ala ala ala
901/301                               931/311
gtt gcc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc
ala leu met met met met met met met met met met met met met met met met met

```

82/185

1021/341  
 cgc acg ctg gtc ggg ccg atg cgg gta ctg cgt gat ggg gcg ctc aag gtt gct cat acc  
 arg thr leu val gly pro met arg val leu arg asp gly ala leu lys val ala his thr  
 1081/361  
 gat ctc gac gcc gag atc gcg gcg gtc cgc gcc gcc gac gag ccg atc ccc gag cca ctg  
 asp leu asp gly glu ile ala ala val arg ala gly asp glu pro ile pro glu pro leu  
 1141/381  
 gcg gtg tac acc acc gag gaa atc ggt cag gtc gcg cat gcg gtc gac gag ctg cac acc  
 ala val tyr thr thr glu glu ile gly gln val ala his ala val asp glu leu his thr  
 1201/401  
 cgg gcc ctg ttg ctg gcc gcc gag gaa acg cgg ttg cga ctg ctg gtc aac gag atg ttt  
 arg ala leu leu leu ala gly glu glu thr arg leu arg leu leu val asn glu met phe  
 1261/421  
 gag acc atg tgg cgg cgt agc cgt tcc ctg gtc gac cag cag ctg tgg gtc atc gac caa  
 glu thr met ser arg arg ser arg ser leu val asp gln gln leu ser val ile asp gln  
 1321/441  
 ctg gag cgc aac gag gag gat ccc gcc cga ctc gac agc ett ttc cgg ctc gat cat ctg  
 leu glu arg asn glu glu asp pro ala arg leu asp ser leu phe arg leu asp his leu  
 1381/461  
 gcc gcc cgg ctg cgc cgc aac agc gcc aac ctg ctg gtg ctg gcc ggt gcg cag att acc  
 ala ala arg leu arg arg asn ser ala asn leu leu val leu ala gly ala gln ile thr  
 1441/481  
 cgt gac cac cgc gag ccg gtg ccg ctg tca acc gtg atc agc gcc gcc gtg tca gag gtc  
 arg asp his arg glu pro val pro leu ser thr val ile ser ala ala val ser glu val  
 1501/501  
 gag gac tat cgc cgc gtc gac atc gcg agg gta ccc gac tgt gcg gta gtc gcc gca gcg  
 glu asp tyr arg arg val asp ile ala arg val pro asp cys ala val val gly ala ala  
 1561/521  
 gct ggt gcc gtc att cat ctg ctt gcc gag ctg atc gac aac gcg ttg cgc tac tgg tca  
 ala gly gly val ile his leu leu ala glu leu ile asp asn ala leu arg tyr ser ser  
 1621/541  
 ccg acc acc ccc gth cgg gtc gcc gcc gca atc gcc agc gaa gcc agt gtt atg ctg cga  
 pro thr thr pro val arg val ala ala ala ile gly ser glu gly ser val leu leu arg  
 1681/561  
 atc tgg gat tcc gcc ctg gcc atg acc gat gct gat cgg cgg atg gcc aat atg cgg ctg  
 ile ser asp ser gly leu gly met thr asp ala asp arg arg met ala asn met arg leu  
 1741/581  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg ala gly gly glu val thr pro asp ser ala arg his met gly leu pro val val gly  
 1801/601  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 1861/621  
 gcc acc gcc acc acc gcc gag gtc tac ctc arg gta gtc gtc gtc gtc gtc gtc gtc gtc  
 gly thr gly thr thr ala glu val tyr leu pro leu ala val leu glu gly thr ala arg  
 1921/641  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 1981/661  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2041/681  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2101/701  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2161/721  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2221/741  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2281/761  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2341/781  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2401/801  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2461/821  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2521/841  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2581/861  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2641/881  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2701/901  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2761/921  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2821/941  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2881/961  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 2941/981  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly  
 3001/1001  
 gag gct gct gct gag gtc acc ccc gat arg arg gct gct gct gct gct gct gct gct gct  
 arg leu ala gly arg his gly ile arg val gly leu arg gly pro val thr gly glu gly

FIGURE 11 (continued)

SEQUENCE LISTING (continued)

83/185

2041/681  
 cgc cgt acc ccg ggg tcc agt ggc atc gcc gac gtc ccg gcc cag ccg atg cag cag ccg  
 arg arg thr pro gly ser ser gly ile ala asp val pro ala gln pro met gln gln arg  
 2101/701  
 ccg cgc gag ctg aaa aca ccc tgg tgg gag gat agg ttt caa cag gag ccc aaa caa ccg  
 arg arg glu leu lys thr pro trp trp glu asp arg phe gln gln glu pro lys gln pro  
 2161/721  
 ccc gca cca gaa ccg cga ccg gcg ccg ccg ccc gcc aaa ccc gcg cca ccg gcg gcc ccg  
 pro ala pro glu pro arg pro ala pro pro pro ala lvs pro ala pro pro ala gly pro  
 2221/741  
 gtt gat gac gac gtc atc tac ccg ccg atg ctc tcc gag atg qtg ggt gac ccg cac gag  
 val asp asp asp val ile tyr arg arg met leu ser glu met val gly asp pro his glu  
 2281/761  
 ctg gcc cac agc ccc gat ctg gac tgg aag tgg gtg tgg gac cac gcc tgg tgg gcg gcc  
 leu ala his ser pro asp leu asp trp lys ser val trp asp his gly trp ser ala ala  
 2341/781  
 acc gag gcc gcg gac aag ccc gtg cag tcc ccc ccg gac tac gcc ctg ccg qtg ccg gaa  
 ala glu ala ala asp lys pro val gln ser arg thr asp tyr gly leu pro val arg glu  
 2401/801  
 ccc ggg gcc ccg tta gtg ccg ggg gcg gcg gtg ccc aag gaa ccc cat ccg gag cat ccg  
 pro gly ala arg leu val pro gly ala ala val pro glu gly pro asp arg glu his pro  
 2461/821  
 ggt gca gcg cta gca tcc aac gcc gga ctt cat ccc gcc cga gcg ccg ccg cac gcg gct  
 gly ala ala leu ala ser asn gly gly leu his pro gly arg ala pro arg his ala ala  
 2521/841  
 gcg gta cgc gac ccc gac gcg gtt cgt gcc tcc atc agc agc cat ttc gcc gcc gtg ccg  
 ala val arg asp pro asp ala val arg ala ser ile ser ser his phe gly gly val arg  
 2581/861  
 acc ggg ccg tgg cat gcc cgc gag agc agt ccc gga ccc atc agc cca tga  
 thr gly arg ser his ala arg glu ser ser gln gly pro pro gln gln opa

SEQ ID N° 21D (suite)

FIGURE 11A (suite)

84/185

ORF d'après par Cole et al. (Nature 393:537-544) et contenant Rv336Sc

```

1/1                                31/11
taa ggg tgc ggc cgg tgg cac ggc cgc ggc cac gtg acc atg ttc gcc cgc ccg acc atc
OCH gly cys gly arg trp his gly arg gly his val thr met phe ala arg pro thr ile
61/21                                91/31
ccg gtc gcg gcg gcc gct tct gat att tcc gcc ccg gct caa ccg gcc cgc ggc aaa cct
pro val ala ala ala ala ser asp ile ser ala pro ala gln pro ala arg gly lys pro
121/41                                151/51
cag caa cgc ccg ccg tcc tgg tgg ccg cgc aac tgg ccg gtc cga tgg aaa gtg ttc acg
gln gln arg pro pro ser trp ser pro arg asn trp pro val arg trp lys val phe thr
181/61                                211/71
atc gag ctt ctg ccg ctg gta gtg gcg atg gtg tta gca gga ttg cgg gtc gag gct gag
ile ala leu leu pro leu val val ala met val leu ala gly leu arg val glu ala ala.
241/81                                271/91
atg gcc agc acc agc ggc ctg ccg ctg gtc gcc ccg cgc gcc gaa atg ata ccc gcg atc
met ala ser thr ser gly leu arg leu val ala ala arg ala glu met ile pro ala ile
301/101                               331/111
acg aaa tac atg tgg gcg ctg gaa gtc gcc gtg ctg gcc agc tgg acc gga cac gat gtg
thr lys tyr met ser ala leu asp val ala val leu ala ser ser thr gly his asp val
361/121                               391/131
gag ggg gcg cag aaa aac ttc acc gcc cgc agc tac gag ctg cag acg cga ctg gcc gac
glu gly ala gln lys asn phe thr ala arg lys tyr glu leu gln thr arg leu ala asp
421/141                               451/151
acc gac gtc atc gca gac gtg ccg tgg gga gtg aac agc ctg ctc aac gcc ggt cag gcg
thr asp val ile ala asp val arg ser gly val asn thr leu leu asn gly gly gln ala
481/161                               511/171
ctg ctg gat aag gtg ctg gcc gac agc atc gcc ttg ccg gat ccg gtc acc gcc tac gcg
leu leu asp lys val leu ala asp ser ile gly leu arg asp arg val thr ala tyr ala
541/181                               571/191
ccg ctg ctg ttg acg gcc cag aac gtg att gag ggg tgg ctg ccg gtt gac agc gag caa
pro leu leu leu thr ala gln asn val ile asp ala ser val arg val asp ser glu gln
601/201                               631/211
atc cga acc cag gtg cag ggt ttg agc cga gcc gtt gtt gtt ggt gac atg acc atc
ile arg thr gln val gln gly leu ser arg ala val gly ala arg gly gln met thr met
661/221                               691/231
cag gag atc ctg gtg act cgc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
gln gln ile leu val thr arg gly ala asp gcc ala gln thr gln leu arg gln gln gln
721/241                               751/251
ctt acc atg ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc
val thr leu ala gly thr ala pro ser thr leu phe gly met leu ala ala leu gly ala
781/261                               811/271
ggc tgg agc gac acc aac aac gtc gac gac gac gac gac gac gac gac gac gac gac gac
gly ser gln asp thr lys asp leu gln gln gln met val thr arg met ala ile met ser
841/281                               871/291
atc cgc gcc atc gca ctg gtc atc gac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
asp pro ala thr ala leu val asp gln gln gln gln gln gln gln gln gln gln gln gln

```

901/301  
gac att gcc gag cag gtg atc acc gac acc acc gag gcg gtg acg aag tcg gtg caa agc  
asp ile ala glu gln val ile thr asp thr thr glu ala val thr lys ser val gln ser  
961/321  
cag gcc acc gac cgg cgg gat gcc gcg att cgc gac gcc gtg ctg gtg ttg gcc gcc atc  
gln ala thr asp arg arg asp ala ala ile arg asp ala val leu val leu ala ala ile  
1021/341  
gcg acc gcg atc gtc gtc gtg ttg gtg gtg gcg cgc acg ctg gtc ggg cgg atg cgg gta  
ala thr ala ile val val val leu val val ala arg thr leu val gly pro met arg val  
1081/361  
ctg cgt gat ggg gcg ctc aag gtt gct cat acc gat ctc gac gcc gag atc gcc gcg gtc  
leu arg asp gly ala leu lys val ala his thr asp leu asp gly glu ile ala ala val  
1141/381  
cgc gcc gcc gac gag cgg atc ccc gag cca ctg gcg gtg tac acc acc gag gaa atc ggt  
arg ala gly asp glu pro ile pro glu pro leu ala val tyr thr thr glu glu ile gly  
1201/401  
cag gtc gcg cat gcg gtc gac gag ctg cac acc gcg gcc ctg ttg ctg gcc gcc gag gaa  
gln val ala his ala val asp glu leu his thr arg ala leu leu leu ala gly glu glu  
1261/421  
acg cgg ttg cga ctg ctg gtc aac gag atg ttt gag acc atg tcg cgg cgt agc cgt tcc  
thr arg leu arg leu leu val asn glu met phe glu thr met ser arg arg ser arg ser  
1321/441  
ctg gtc gac cag cag ctg tcg gtc atc gac caa ctg gag cgc aac gag gag gat ccc gcc  
leu val asp gln gln leu ser val ile asp gln leu glu arg asn glu glu asp pro ala  
1381/461  
cga ctc gac agc ctt ttc cgg ctc gat cac ctg gcc gcc cgg ctg cgc cgc aac agc gcc  
arg leu asp ser leu phe arg leu asp his leu ala ala arg leu arg arg asn ser ala  
1441/481  
aac ctg ctg gtg ctg gcc ggt gcg gag att acc cgt gac cac cgc gag cgg gtg cgg ctg  
asn leu leu val leu ala gly ala gln ile thr arg asp his arg glu pro val pro leu  
1501/501  
tca acc gtg atc agc gcc gcc gag tta gag gtc gag gcc tat cgc cgc gtc gac atc gag  
ser thr val ile ser ala ala val ser glu val glu asp tyr arg arg val asp ile ala  
1561/521  
agg gta ccc gac tgt gcg gta gtc gcc gcc gcg gat ggt gcc gtc att cat ctg ctt ggt  
arg val pro asp cys ala val val gly ala ala ala gly gly val ile his leu leu ala  
1621/541  
ag ctg atc acc acc gcg tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta tta  
gln leu ile asp asn ala leu arg tyr ser ser pro thr thr pro val val val ala ala  
1681/561  
tta acc gcc gta gaa gcc agt gta ctg ctg acc atc tta gat tta gcc ctg gta atc a  
ala ile gly ser ala gly ser val leu leu ala ala acc agc ser arg leu gly met thr  
1741/581  
tat gcc gat gta gga atc gcc cat atc gcc ctg gag tta tta gta tta tta acc acc acc  
asp ala acc gta arg met ala asp met arg leu arg ala gly gly ala val thr pro arg  
1801/601  
tct tta  
ala ala met gly leu thr ala val gly ser ala gly gly gta tta tta tta tta tta tta tta

86/185

1861/621 1891/631  
 gtc ggg ctg cgc ggt ccg gtg acc ggt gaa cag ggc acc ggc acc acc gcc gag gtc tac  
 val gly leu arg gly pro val thr gly glu gln gly thr gly thr thr ala glu val tyr  
 1921/641 1951/651  
 ctg ccg cta gcc gtg ctc gag ggg acg gcc cca gcg cag ccg cca aag ccg cgg gta ttt  
 leu pro leu ala val leu glu gly thr ala pro ala gln pro pro lys pro arg val phe  
 1981/661 2011/671  
 gcg atc aag ccg ccg tgt cct gaa ccc gcg gcg gcc gat ccg acg gac gtt ccc gcc gcc  
 ala ile lys pro pro cys pro glu pro ala ala ala asp pro thr asp val pro ala ala  
 2041/681 2071/691  
 atc ggg ccg cta cca ccg gtc acg ttg ctc ccg cgc cgt acc ccg ggg tcc agt gcc atc  
 ile gly pro leu pro pro val thr leu leu pro arg arg thr pro gly ser ser gly ile  
 2101/701 2131/711  
 gcc gac gtc ccg gcc cag ccg atg cag cag cgg cgg cgc gag ctg aaa aca ccc tgg tgg  
 ala asp val pro ala gln pro met gln gln arg arg arg glu leu lys thr pro trp trp  
 2161/721 2191/731  
 gag gat agg ttt caa cag gag ccc aaa caa ccg ccc gca cca gaa ccg cga ccg gcg ccg  
 glu asp arg phe gln gln glu pro lys gln pro pro ala pro glu pro arg pro ala pro  
 2221/741 2251/751  
 ccg ccc gcc aaa ccc gcg cca ccg gcg ggc ccg gtt gat gac gac gtc atc tac cgg cgg  
 pro pro ala lys pro ala pro pro ala gly pro val asp asp asp val ile tyr arg arg  
 2281/761 2311/771  
 atg ctc tcc gag atg gtg ggt gac ccg cac gag ctg gcc cac agc ccc gat ctg gac tgg  
 met leu ser glu met val gly asp pro his glu leu ala his ser pro asp leu asp trp  
 2341/781 2371/791  
 aag tcc gtg tgg gac cac ggc tgg tcc gcg gcc gcc gag gcc gcg gac aag ccc gtg cag  
 lys ser val trp asp his gly trp ser ala ala ala glu ala ala asp lys pro val gln  
 2401/801 2431/811  
 tcc cgc acg gac tac ggc ctg ccg gtg cgc gaa ccc ggg gcc cgg tta gtg ccg ggg gag  
 ser arg thr asp tyr gly leu pro val arg glu pro gly ala arg leu val pro gly ala  
 2461/821 2491/831  
 gcg gtg ccc gag gga ccc gat cgg gag cat ccg agt gca gag cta gcc tcc aac gcc gcc  
 ala val pro glu gly pro asp arg glu his pro gly ala ala leu ala ser asp gly gly  
 2521/841 2551/851  
 ctt cat ccc gcc cga gcg ccg cgg cac gcg gcc gcc gta cga gac ccc gac gcg gtt cgt  
 leu his pro gly arg ala pro arg his ala ala ala val arg asp pro asp ala val arg  
 2581/861 2611/871  
 tcc tcc atc arg arg cat ttc arg gac gtc ccc arg gac gtc ccc arg gac gtc ccc arg gac gtc  
 ala ser ile ser ser his phe gly gly val arg tcc arg arg ser his ala arg ala arg  
 2641/881 2671/891  
 tcc tcc atc arg arg cat ttc arg gac gtc ccc arg gac gtc ccc arg gac gtc ccc arg gac gtc  
 ser ala gly pro asp ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala

SEQ ID NO: 215 (continued)

87/185

31/11  
 CTA CGA CAA GGC AAA GGA GCA CAG GGT GAA GCG TGG ACT GAC GGT CGC GGT AGC CGG AGC  
 leu arg gln gly lys gly ala gln gly glu ala trp thr asp gly arg gly ser arg ser  
 61/21  
 CGC CAT TCT GGT CGC AGG TCT TTC CGG ATG TTC AAG CAA CAA GTC GAC TAC AGG AAG CGG  
 arg his ser gly arg arg ser phe arg met phe lys gln gln val asp tyr arg lys arg  
 121/41  
 TGA GAC CAC GAC CGC GGC AGG CAC GAC GGC AAG CCC CGG CGC CGC ATC CGG GCC GAA GGT  
 OPA asp his asp arg gly arg his asp gly lys pro arg arg arg ile arg ala glu gly  
 181/61  
 CGT CAT CGA CGG TAA GGA CCA GAA CGT CAC CGG GTC TGT GGT GTG CAC AAC CGC GGC CGG  
 arg his arg arg OCH gly pro glu arg his arg val cys gly val his asn arg gly arg  
 241/81  
 CAA TGT CAA CAT CGC GAT CGC CGC GGC GGC GAC CGC CAT TGC CGC CGT GCT CAC CGA CGG  
 gln cys gln his arg asp arg arg gly gly asp arg his cys arg arg ala his arg arg  
 301/101  
 CAA CCC TCC GGA GGT GAA GTC CGT TGG GGT CGG TAA CGT CAA CGG CGT CAC GGT GGG ATA  
 gln pro ser gly gly glu val arg trp ala arg OCH arg gln arg arg his ala gly ile  
 361/121  
 CAC GTC GGG CAC CGG ACA GGG TAA CGC TCG GGA ACT AAG CAC GGC AGC CAC TAC AAG ATC  
 his val gly his arg thr gly OCH arg ser ala thr lys asp gly ser his tyr lys ile  
 391/151

SEQ ID N° 22A

FIGURE 22A

31/11  
 TAC GAC AAG GCA AAG GAG CAC AGG GTC AAG CTA GAA TTA AAT GTC GAG GTA GCG GGA GGC  
 tyr asp lys ala lys glu his arg val lys arg gly leu thr val ala val ala gly ala  
 62/21  
 GCC ATT CTG CTC GCA GGT GGT TCG GGA TGT TTA AAT AAT TGG AAT ACA GGA AGC GGT  
 ala ile leu val ala gly leu ser gly ty arg ser arg ty arg thr thr gly ser gly  
 122/41  
 GAG ACC AAG AAT GCG GCA GAT ATG ATT TTA ACC GAT GAT GAA TTA TTA GTC GTC AAT GT  
 glu thr thr thr ala ala gly thr thr ala leu arg arg ala ala arg gly pro lys val  
 182/61  
 GTC ATT GAC GAT AAG GAT GAT AAT GTT AAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 val ile arg thr lys asp thr thr val thr thr thr thr thr thr thr thr thr thr thr  
 242/81  
 AAT GTT AAT AAT GAG ATT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 asn val leu ile ala ile gly gly ala ala thr thr thr thr thr thr thr thr thr thr  
 302/101  
 AAT GGT GAT GAT GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 asn gln thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
 362/121  
 AAT TTA GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT  
 asn thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

88/185

33/11  
 ACG ACA AGG CAA AGG AGC ACA GGG TGA ACC GTG GAC TGA CGG TCG CGG TAG CCG GAG CCG  
 thr thr arg gln arg ser thr gly OPA ser val asp OPA arg ser arg AMB pro glu pro  
 63/21  
 CCA TTC TGG TCG CAG GTC TTT CCG GAT GTT CAA GCA ACA AGT CGA CTA CAG GAA GCG GTG  
 pro phe trp ser gln val phe pro asp val gln ala thr ser arg leu gln glu ala val  
 123/41  
 AGA CCA CGA CCG CGG CAG GCA CGA CGG CAA GCG CCG GCG CCG CAT CCG GGC CGA AGG TCG  
 arg pro arg pro arg gln ala arg arg gln ala pro ala pro his pro gly arg arg ser  
 193/61  
 TCA TCG ACG GTA AGG ACC AGA ACG TCA CCG GGT CTG TGG TGT GCA CAA CCG CGG CCG GCA  
 ser ser thr val arg thr arg thr ser pro gly leu tip cys ala gln pro arg pro ala  
 243/81  
 ATG TCA ACA TCG CGA TCG GCG GGG CCG CGA CCG GCA TTG CCG CCG TGC TCA CCG ACG GCA  
 met ser thr ser arg ser ala gly arg arg pro ala leu pro pro cys ser pro thr ala  
 303/101  
 ACC CTC CCG AGG TGA AGT CCG TTG GGC TCG GTA ACG TCA ACG GCG TCA CCG TGG GAT ACA  
 thr leu arg arg OPA ser pro leu gly ser val thr ser thr ala ser arg tip asp thr  
 363/121  
 CGT CCG GCA CCG GAC AGG GTA ACG CTC GCG AAG CAA GGA CCG CAG CCA CTA CAA GAT C  
 arg arg ala pro asp arg val thr leu gly asp gln gly arg gln pro leu gln asp

SEQ ID N° 22C

FIGURE 22C

51/11  
 GCA CAA CCG CGG CCG GCA ATG TCA ACA TCG GCA TGG GCG GCG CGG CGA CCG GCA TTG CCG  
 ala gln pro arg pro ala met ser thr ser arg ser ala gly arg arg pro ala leu pro  
 61/21  
 CCG TGC TCA CCG ACG GCA ACG CTC CCG AGT TGA ACT CCG TTG GCG TGG GTA ACG TCA ACG  
 pro cys ser pro thr ala thr leu arg arg OPA ser pro leu gly ser val thr ser thr  
 121/41  
 GCG TCA CCG TGG GAT ACA CGT CCG GCA CCG GAC AGG GTA ACG CTC CCG CAA CCA AGG ACG  
 ala ser arg trp asp thr arg arg ala pro asp arg val thr pro arg gln pro arg met  
 181/61  
 GCA GCG ATT ACA AGA TCA CAG GGT GAA GCG TGA AGT GAT GAT GAT GAT GAT GAT GAT GAT  
 ala ala thr thr arg ser gln gly gln ala tip thr asp gly asp gly ser arg ser arg  
 241/81  
 CAT TGT GGT CCG AGG TGT TCG CCG ATG TTT AAG TAA TAA TTT TAT TAT AGG AAA TAT TAA  
 his ser gly arg arg ser phe arg met thr thr ala thr thr thr thr thr thr thr thr  
 301/101  
 GAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT  
 asp his asp arg gly arg his asp gly lys pro arg arg arg arg arg arg arg arg arg  
 361/121  
 ATT GAT GAT AAG GAT TAG AAA GGT AAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT  
 ile asp gly lys asp ala ala val thr gly ser val asp thr thr thr thr thr thr thr  
 421/141  
 GTT AAT AAT TAT AAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT  
 val met ile ile ile gly arg arg arg arg arg arg arg arg arg arg arg arg arg arg

SEQ ID N° 22A

SEQUENCE ALIGNMENT (FIGURE 25)



89/185

32/11  
 CAC AAC CGC GGC CGG CAA TGT CAA CAT CGC GAT CGG CGG GGC GGC GAC CGG CAT TGC CGC  
 his asn arg gly arg gln cys gln his arg asp arg arg gly gly asp arg his cys arg  
 62/21  
 CGT GCT CAC CGA CGG CAA CCC TCC GGA GGT GAA GTC CGT TGG GCT CGG TAA CGT CAA CGG  
 arg ala his arg arg gln pro ser gly gly glu val arg trp ala arg OCH arg gln arg  
 122/41  
 CGT CAC GCT GGG ATA CAC GTC GGG CAC CGG ACA GGG TAA CCC CTC GGC AAC CAA GGA CGG  
 arg his ala gly ile his val gly his arg thr gly OCH arg leu gly asn gln gly arg  
 182/61  
 CAG CCA CTA CAA GAT CAC AGS GTG AAG COT GGA CTG ACG GTC GCG GTA GCC GGA GGC GCC  
 gln pro leu gln asp his arg val lys arg gly leu thr val ala val ala gly ala ala  
 242/81  
 ATT CTG GTC GCA GGT CIT TCG GGA TGT TCA AGC AAC AAG TCG ACT ACA GGA AGC GGT GAG  
 ile leu val ala gly leu ser gly cys ser ser asn lys ser thr thr gly ser gly glu  
 302/101  
 ACC ACG ACC GCG GCA GGC ACG ACG GCA AGC COT GGC GCG GCT CCG GGC CGA AGG TCG TCA  
 thr thr thr ala ala gly thr thr ala ser pro gly ala ala pro gly arg arg ser ser  
 362/121  
 TCG ACG GTA AGG ACC ASA ACG TCA CCG GCT CCG TGG TGT GCA CAA CCG CGG CCG SCA ATG  
 ser thr val arg thr arg thr ser pro ala pro trp cys ala gln pro arg pro ala met  
 422/141  
 TCA ACA TCG CGA TCG SCG GGG CGG CGA CCG GCA TTG CCG CCG TGC TCA CCG ACG GCA ACC  
 ser thr ser arg ser ala gly arg arg pro ala leu pro pro cys ser pro thr ala thr  
 482/161  
 CTC CGG AGG TGA AGT CCG TTG GGC TCG GTA ACG TCA ACG GCG TCA CCG TGG GAT ACA CGT  
 leu arg arg OPA ser pro leu gly ser val thr ser thr ala ser arg trp asp thr arg  
 542/181  
 CGG GCA CCG GAC AGG GTA ACG COT CCG CAA GGA ACG ACG GCA GGC ACT ACA AGA TC  
 arg ala pro asp arg val thr pro arg gln pro arg thr ala ala thr thr arg

SEQ ID N° 23B

FIGURE 23B

90/185

33/11  
 ACA ACC GCG GCC GGC AAT GTC AAC ATC GCG ATC GGC GGG GCG GCG ACC GGC ATT GCC GCC  
 thr thr ala ala gly asn val asn ile ala ile gly gly ala ala thr gly ile ala ala  
 63/21  
 93/31  
 GTG CTC ACC GAC GGC AAC CCT CCG GAG GTG AAG TCC GTT GGG CTC GGT AAC GTC AAC GGC  
 val leu thr asp gly asn pro pro glu val lys ser val gly leu gly asn val asn gly  
 123/41  
 153/51  
 GTC ACG CTG GGA TAC ACG TCG GCC ACC GGA CAG GGT AAC GCC TCG GCA ACC AAG GAC GGC  
 val thr leu gly tyr thr ser gly thr gly gln gly asn ala ser ala thr lys asp gly  
 183/61  
 213/71  
 AGC CAC TAC AAG ATC ACA GGG TGA AGC GTG GAC TGA CCG TCG CGG TAG CCG CAG CCG CCA  
 ser his tyr lys ile thr gly opa ser val asp opa arg ser arg AMB pro glu pro pro  
 243/81  
 273/91  
 TTC TGG TCG CAG GTC TTT CCG GAT GTT CAA SCA ACA AGT CGA CTA CAG GAA GCG GTG AGA  
 phe trp ser gln val phe pro asp val gln ala thr ser arg leu gln glu ala val arg  
 303/101  
 333/111  
 CCA CGA CCG CGG CAG GCA CGA CCG CAA GCG CCG GCG CCG CTC CGG GCC GAA GGT CGT CAT  
 pro arg pro arg gln ala arg arg gln ala pro ala pro leu arg ala glu gly arg his  
 363/121  
 393/131  
 CGA CCG TAA GGA CCA GAA CGT CAC CCG CTC CGT GGT GTS CAC AAC CGC GGC CCG CAA TGT  
 arg arg OCH gly pro glu arg his arg leu arg gly val his asn arg gly arg gln cys  
 423/141  
 453/151  
 CAA CAT CCG GAT CGG CCG GGC GGC GAC CCG CAT TGC CGC CGT GGT CAC CGA CCG CAA CCC  
 gln his arg asp arg arg gly gly asp arg his cys arg arg ala his arg arg gln pro  
 483/161  
 513/171  
 TCC GGA GGT GAA GTC CGT TGG CCT CGG TAA GGT CAA CCG CGT CAC GCT GGG ATA CAC GTC  
 ser gly gly glu val arg trp ala arg OCH arg gln arg arg his ala gly ile his val  
 543/181  
 573/191  
 GGG CAC CCG ACA GGG TAA CCG CTC GGC AAC CAA GGA CCG CAG CCA CTA CAA GAT C  
 gly his arg thr gly OCH arg leu gly asn gln gly arg gln pro leu gln asp

SEQ ID NO 23C

FIGURE 23C



92/185

33/11  
 AAC GAC AGG CAA AGG AGC ACA GGG TGA AGC GTG GAC TGA CCG TCG CCG TAG CCG GAG CCG  
 asn asp arg gln arg ser thr gly OPA ser val asp OPA arg ser arg AMB pro glu pro  
 63/21  
 93/31  
 CCA TTC TGG TCG CAG GTC TTT CCG GAT GTT CAA GCA ACA AGT CGA CTA CAG GAA GCG GTG  
 pro phe trp ser gln val phe pro asp val gln ala thr ser arg leu gln glu ala val  
 123/41  
 153/51  
 AGA CCA CGA CCG CCG CAG GCA CGA CCG CAA GCG CCG GCG CCG CTC CCG GCC GAA GGT CGT  
 arg pro arg pro arg gln ala arg arg gln ala pro ala pro leu arg ala glu gly arg  
 183/61  
 213/71  
 CAT CGA CCG TAA GGA CCA GAA CGT CAC CCG CTC CGT GGT GTG CAC AAC CCG GGC CCG CAA  
 his arg arg OCH gly pro glu arg his arg leu arg gly val his asn arg gly arg gln  
 243/81  
 273/91  
 TGT CAA CAT CSC GAT CCG CCG GGC GGC GAC CCG CAT TGC CCG CGT GGT CAC CGA CCG CAA  
 cys gln his arg asp arg arg gly gly asp arg his cys arg arg ala his arg arg gln  
 303/101  
 333/111  
 CCG TCG GGA GGT GAA GTC CCG TGG GGT CCG TAA GGT CAA GGT CCG CAC GGT GCG ATA CAG  
 pro ser gly gly glu val arg trp ala arg OCH arg gln arg arg his ala gly ile his  
 363/121  
 393/131  
 GTC GGG CAC CCG ACA GGG TAA CCG CTC GCG AAC CAG GGA CTA CAG CCA CTA CAA GAT C  
 val gly his arg thr gly OCH arg leu gly asn gln gly arg gln pro leu gln asp

SEQ ID N° 24C

FIGURE 24C

Amoeba proteus

100% amino acid identity

SEQ ID N° 25

FIGURE 25

Amoeba proteus

100% amino acid identity

100% amino acid identity

31/11

CCT ACC AGC AAG AGC CCA GGG CTT CAC AGG ACC TAA AAG GAG TAG CGC CCA TGG GCT TGA  
 pro thr ser lys ser pro gly leu his arg thr OCH lys glu AMB arg pro trp ala OPA  
 61/21 91/31  
 TCC AAT TTT CCT TCC GCC CCG TGC AAT ACC ATC TGC AAG ACC AGC GAC GGC CCG TGG TTG  
 ser asn phe pro ser ala pro cys asn thr ile cys lys thr ser asp gly pro trp leu  
 121/41 151/51  
 CGG TCG CCC AGC TTG CGG AAA CGG GGT ATG GAC CCT GCC GTA CCG TTG TTG CCA CTT GAT  
 arg ser arg ser leu arg lys arg gly met asp pro ala val pro leu leu pro leu asp  
 181/61 211/71  
 GTG CTC GCT CTC CAC CCG TCG GGG GGC GAA ACC CAT TCC GAC ACT GGG ATC CTC AAA ACC  
 val val ala leu his pro ser gly gly glu ser his ser asp thr gly ile leu lys thr  
 241/81 271/91  
 TCG GCT GAG TGT CTG CAG GGC TCC GCG GAG CAG CCG ATC ATC ACC ATG TAC GAA CTG AAT  
 ser ala glu cys leu glu gly ser gly glu gln pro ile ile thr met tyr glu leu asn  
 301/101 331/111  
 AAG TCC CCC CCG CGC GAC TTC CAG ACA TTT GTT GTG GTT TCG GTT GAG GCC GAG GCG AGG  
 lys ser pro pro arg asp phe gln thr phe val val val ser val glu ala glu ala arg  
 361/121 391/131  
 CTG ATT TCG CAG CAA GGC CTC TCC GCG TCG CAG CAT CCT TCG GCG GAT CGC GCG GGA GTG  
 leu ile ser gln gln ala val ser gly ser gln his arg cys gly asp arg gly ala val  
 421/141  
 GTC GGA CGA GTC GTC GTC AAC GAC CAC GAT C  
 val gly arg val val val asn asp his asp

SEQ ID N° 27A

FIGURE 27A

33/11

TAC CAG CAA GAG CCC AGG GCT TCA CAG GAC CTA AAA GGA GTA GCG CCC ATG GGC TTG ATC  
tyr gln gln glu pro arg ala ser gln asp leu lys gly val ala pro met gly leu ile  
63/21 93/31

CAA TTT TCC TTC CGC CCC GTG CAA TAC CAT CTG CAA GAC CAG CGA CGG CCC GTG GTT GCG  
gln phe ser phe arg pro val gln tyr his leu gln asp gln arg arg pro val val ala  
123/41 153/51

GTC GCG CAG CTT GCG GAA ACG GGG TAT GGA CCC TGC CGT ACC GTT GTT GCC ACT TGA TGT  
val ala gln leu ala glu thr gly tyr gly pro cys arg thr val val ala thr OPA cys  
183/61 213/71

CST CGC TCT CCA CCC GTC GGG GGG CGA AAG CCA TTC CCA CAC TGG GAT CCT CAA AAC GTC  
arg arg ser pro pro val gly gly arg lys pro phe arg his trp asp pro gln asp val  
243/81 273/91

GGC TGA GTG TGT GCA GGG CTC CGG GGA GCA GCC GAT CAT CAC CAT GTA CGA ACT GAA TAA  
gly OPA val ser ala gly leu arg gly ala ala asp his his his val arg thr glu OPA  
303/101 333/111

GTC CCC CCC GCG CGA CTT CCA CAC ATT TGT TGT GGT TTC GGT TGA GGC CGA GGC GAG GCT  
val pro pro ala arg leu pro asp ile cys cys gly phe gly OPA gly arg gly glu ala .  
363/121 393/131

CAT TTC GCA GCA AGC GGT CTC CGG GTC GCA GCA TCG TTS CGG CGA TCG CGG CGC AGT CCG  
his phe ala ala ser gly leu arg val ala ala ser leu arg arg ser arg arg ser arg  
423/141

CGG ACG AGT CCG CCG CAA CGA CGA CGA TC  
arg thr ser arg arg gln arg pro arg

FIGURE 27C

NYTSHVATNTRGDIHVDIALALPAAVALLABEJATGASEPQACEEVAN  
TUGSHAKKMOOYLKHPFTYOLZTAVLEJLHNSHUALKAFHWANPM  
VAGLHIALBIPITLINTFOLPITLQJALHNMNYIYABE

FIGURE 28

MGKQLAALAAALVGACMLAAGCTNVVDGTAVAADKSGPLHQDPIPVETSALEGLLLDLSQINAALGATS  
MKVWFNAKAMWDWSKSVADKNCILADGPAQEKVYAGTGFTWTAMRGQRLLDSDIDSKKRDRHYAIQAVV  
GFPTAHDAAEFYSSSVQSWSSCSNRRFVEVTFTPGGDDAAWTVADVVNDNGMLSSSQVQEGGDGWTCC  
RALTARNNVTDIDVTCAYSQPDLVETAIGIANQIAAKVAKQ

FIGURE 30

FIGURE 31A





97/185

1/1	31/11
aga ctg gtg tac acg gag acc aag ctg aac	tcg gca ttc tcc ttc ggc ggg cct aag tgt
arg leu val tyr thr glu thr lys leu asn	ser ala phe ser phe gly gly pro lys cys
61/21	91/31
cta gtg aag gtc att cag aaa ctg tcg ggc	ttg agc atc aac cgg ttc atc gcg att gac
leu val lys val ile gln lys leu ser gly	leu ser ile asn arg phe ile ala ile asp
121/41	151/51
ttc gtc ggt ttc gcg cgg atg gtc gag gcc	ctc ggc ggc gtc gag gta tgc agc acc acc
phe val gly phe ala arg met val glu ala	leu gly gly val glu val cys ser thr thr
181/61	211/71
cgg ttg cgg gac tac gaa ctg ggc acg gtg	ctg gag cag gcc gga cgc cag gtc att gac
pro leu arg asp tyr glu leu gly thr val	leu glu his ala gly arg gln val ile asp
241/81	271/91
ggg ccg acc gcg ctg aac tat gtg cgc gcc	cgc cag gtc acc acc gag agc aat ggc gac
gly pro thr ala leu asn tyr val arg ala	arg gln val thr thr glu ser asn gly asp
301/101	331/111
tac ggg cgc atc aaa cgc cag cag ttg ttt	ttg tcc tcc ctg ctg cgt tcc atg atc
tyr gly arg ile lys arg gln gln leu phe	leu ser ser leu leu arg ser met ile

SEQ ID N° 32A

FIGURE 32A

[illegible]

98/185

1/1  
act ggt gta cac gga gac caa ggt gaa ctc ggc att ctc ctt cgg cgg gcc taa gtg tct  
thr gly val his gly asp gln ala glu leu gly ile leu leu arg arg ala OCH val ser  
61/21  
agt gaa ggt cat tca gaa act gtc ggg ctt gag cat caa cgg gtt cat cgc gat tga ctt  
ser glu gly his ser glu thr val gly leu glu his gln pro val his arg asp OPA leu  
121/41  
cgt cgg ttt cgc gcg gat ggt cga ggc cct cgg cgg cgt cga ggt atg cag cac cac ccc  
arg arg phe arg ala asp gly arg gly pro arg arg arg arg gly met gln his his pro  
181/61  
gtt gcg gga cta cga act ggg cac ggt gct gga gca cgc cgg aag cca ggt cat tga cgg  
val ala gly leu arg thr gly his gly ala gly ala arg arg thr pro gly his OPA arg  
241/81  
gcc cag cgc gct gaa cta tgt gag cgc tgg cca ggt cac cac cga gag caa tgg cga cta  
ala asp arg ala glu leu cys ala arg ser pro gly his his arg glu gln trp arg leu  
301/101  
cgg gcg cat caa aag cca cca gtt gtt ttt gtc gtc gct gct gcg ttc gat gat c  
arg ala his gln thr pro ala val val phe val val ala ala ala phe asp asp

SEQ ID N° 32C

FIGURE 32C

séquence Rv0822c prédite par Cole et al. (Nature 393:537-544) et contenant seq 32A

[illegible]

99/185

[illegible]

100/185

1921/641 1951/651  
 ttc agc gcg gtg cgc gct ccc ctg ccg agt ggc tcc acc gtc agc gtg cag ata agc cgc  
 phe ser ala val arg ala pro leu pro ser gly ser thr val ser val gln ile ser arg  
 1981/661 2011/671  
 aac tcc tcc agc cca ccg acc aag ctg ccc gag gac ctg acg gtc acc aac gcc gcc gac  
 asn ser ser ser pro pro thr lys leu pro glu asp leu thr val thr asn ala ala asp  
 2041/681  
 acc acc tgc gag tag  
 thr thr cys glu AMB

SEQ ID N° 32D (suite 2)  
 FIGURE 32D (suite 2)

ORF d'après Cole et al. (Nature 393:537-544) et contenant Pv2922c

1/1 31/11  
 tag gac atg agt gac ggc gag agc gcc ggg ccg tgg gaa cgg ctc tcc gag tca gaa ttc  
 AMB asp met ser asp gly glu ser ala ala pro trp ala arg leu ser glu ser ala phe  
 61/21 91/31  
 ccc gat ggt gtt gac cga tgg atc acg gta ccg ccc gcc aca tgg atg gaa gcc cag ggt  
 pro asp gly val asp arg trp ile thr val pro pro ala thr trp val ala ala gln gly  
 121/41 151/51  
 ccg cgg gac acc cag aat gtc ggc tgt cat gcc acc gcc gcc gtt agt gtg gcc gat ctg  
 pro arg asp thr gln asn val gly cys his ala thr gly ala val ser val ala asp leu  
 181/61 211/71  
 atc gcc agg ctc gcc ccc gct ttt cct gac ctc ccc agc cag cgt cat gtc gcc ccc gaa  
 ile ala arg leu gly pro ala phe pro asp leu pro thr his arg his val ala pro glu  
 241/81 271/91  
 ccc gag cca tcc gcc cgc gcc ccg aag gtc tac gat gac gcc gac gac cag cag gac acc  
 pro glu pro ser gly arg gly pro lys val his asp asp ala asp asp gln gln asp thr  
 301/101 331/111  
 gag gct atc gcc atc ccg gcc cag tcc ctc cag ttc ctc tgg gat ctc tcc gac ctc cgg  
 gln ala ile ala ile pro ala his ser leu glu gaa ccc ser ala leu pro asp leu arg  
 361/121 391/131  
 gaa gcc aac tat ccg gcc gac gat gac gcc agc ccc gaa gcc gac tta gcc gcc aag gac  
 ala ala asn tyr pro arg ala asp his ala arg arg ala pro ala leu pro gly lys gln  
 421/141 451/151  
 cta acc gaa tgg gct cca atg gcc ctc ttc gac atc ctc gaa ttc ttc ttc ttc ttc ttc  
 leu thr gly ser ala arg val arg pro leu arg ala arg ala thr ser pro ala pro ala  
 481/161 511/171  
 gaa gaa gat gaa aat tcc gaa ccc gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa gaa  
 gaa  
 541/181 571/191  
 gaa  
 gaa  
 601/201 631/211  
 gaa  
 gaa  
 661/221 691/231  
 gaa  
 gaa  
 721/241 751/251  
 gaa  
 gaa  
 781/261 811/271  
 gaa  
 gaa  
 841/281 871/291  
 gaa  
 gaa  
 901/301 931/311  
 gaa  
 gaa  
 961/321 991/331  
 gaa  
 gaa  
 1021/341 1051/351  
 gaa  
 gaa  
 1081/361 1111/371  
 gaa  
 gaa  
 1141/381 1171/391  
 gaa  
 gaa  
 1201/401 1231/411  
 gaa  
 gaa  
 1261/421 1291/431  
 gaa  
 gaa  
 1321/441 1351/451  
 gaa  
 gaa  
 1381/461 1411/471  
 gaa  
 gaa  
 1441/481 1471/491  
 gaa  
 gaa  
 1501/501 1531/511  
 gaa  
 gaa  
 1561/521 1591/531  
 gaa  
 gaa  
 1621/541 1651/551  
 gaa  
 gaa  
 1681/561 1711/571  
 gaa  
 gaa  
 1741/581 1771/591  
 gaa  
 gaa  
 1801/601 1831/611  
 gaa  
 gaa  
 1861/621 1891/631  
 gaa  
 gaa  
 1921/641 1951/651  
 ttc agc gcg gtg cgc gct ccc ctg ccg agt ggc tcc acc gtc agc gtg cag ata agc cgc  
 phe ser ala val arg ala pro leu pro ser gly ser thr val ser val gln ile ser arg  
 1981/661 2011/671  
 aac tcc tcc agc cca ccg acc aag ctg ccc gag gac ctg acg gtc acc aac gcc gcc gac  
 asn ser ser ser pro pro thr lys leu pro glu asp leu thr val thr asn ala ala asp  
 2041/681  
 acc acc tgc gag tag  
 thr thr cys glu AMB

[illegible]

1/1	31/11																		
CGT CAC CTC TGC CAT GGT CCA TCT ACG GTA	TCT GCG ACA AGG GCA GCG TCG ATC CCT CGA																		
arg his leu cys his gly pro ser thr val	ser ala thr arg ala ala ser ile pro arg																		
61/21	91/31																		
CAT GCA GAG TCG GTG TTC GCT TCA CGC GAA	CTA GGC GCG CCT AGC CTG GAC GAG TCC CCG																		
his ala glu ser val phe ala ser arg glu	leu gly ala pro ser leu asp glu ser pro																		
121/41	151/51																		
GGC CGA CAT TCG CCC GAG GCC TTG GCC TCC	ATC ACC TAA TTG TGT GCA AAA CCG TAT CTA																		
gly arg his ser pro glu ala leu ala ser	ile thr OCH leu cys ala lys pro tyr leu																		
181/61	211/71																		
ATT GAT ACG ATT GCG CAC ATG GCT ATC TGG GAT C																			
ile asp thr ile ala his met ala ile trp asp																			

FIGURE 33A

```

1/1
GTC ACC TCT GGC ATG GTC CAT CTA CGG TAT CTS CGA CAA GGG CAG CGT CGA TCG CTC GAC
val thr ser ala met val his leu arg tyr leu arg gln gly gln arg arg ser leu asp
61/21
ATG CAG AGT CGG TGT TCG CTT CAC GCG AAC TAG GGG CGG CTA GCG TGG ACG AGT CCC CGG
met gln ser arg cys ser leu his ala asn AMB ala arg leu ala trp thr ser pro arg
121/41
GCC GAC ATT CGG CCG AGG CCT TGG CCT CGA TCA CTT AAT TGT GTG CAA AAC CGT ATC TAA
ala asp ile arg pro arg pro trp pro pro ser pro asn tyr val ala asn arg ile CCH
181/61
TTG ATA CGA TTG CGG AAG TGG CTA TCT GCG ATC
leu ile arg leu arg thr tyr leu ser gly ile
211/71

```

FIGURE 33B

103/185

séquence Rv1044 prédite par Cole et al. (Nature 393:537-544) et contenant seq33A

```

1/1                                31/11
ttg tgt gca aaa ccg tat cta att gat acg att gcg cac atg gct atc tgg gat cgc ctc
leu cys ala lys pro tyr leu ile asp thr ile ala his met ala ile trp asp arg leu
61/21                                91/31
gtc gag gtt gcc gcc gag caa cat gcc tac gtc acg act cgc gat gcg cga gac atc gcc
val glu val ala ala glu gln his gly tyr val thr thr arg asp ala arg asp ile gly
121/41                                151/51
gtc gac ccc gtg cag ctc cgc ctc cta gcc gcc ggc gga cgt ctt gag cgt gtc gcc cga
val asp pro val gln leu arg leu leu ala gly arg gly arg leu glu arg val gly arg
181/61                                211/71
ggc gtg tac cgg gtg ccc gtg ctg ccg cgt ggt gag cac gac gat ctc gca gcc gca gtc
gly val tyr arg val pro val leu pro arg gly glu his asp asp leu ala ala ala val
241/81                                271/91
tcg tgc act ttg ggc cgt gcc gtt atc tgc cat gag tgg gcc tgg gcg ctt cat gcc ctc
ser trp thr leu gly arg gly val ile ser his ala ser ala leu ala leu his ala leu
301/101                               331/111
gct gac gtg aac ccg tgc cgc atc cat ctc aac gtc cag ccc aac aac cat ccg cgt gcg
ala asp val asn pro ser arg ile his leu thr val pro arg asn asn his pro arg ala
361/121                               391/131
gcc gcc gcc gag ctg tac cga att aac cgc gcc gat ctc cag gaa gcc cac gtc act tgc
ala gly gly glu leu tyr arg val his arg arg asp leu gln ala ala his val thr ser
421/141                               451/151
gtc gac gga ata ccc gtc acg aag gtt gcc gcc aac atc aac gac tgc gtg aag acg gcc
val asp gly ile pro val thr thr val ala arg thr ile lys asp cys val lys thr gly
481/161                               511/171
acg gat cct tat cag ctt cgg gcc gcc atc aac cca gcc gaa gcc gag gcc acg ctt cat
thr asp pro tyr gln leu arg ala ala ile gln arg ala ala ala ala gly thr leu arg
541/181                               571/191
cgt gcc tca gaa gct gag cta cgc gct gcc ctc tgc tgc aac atc gcc gcc tta cgc gct
arg gly ser ala ala glu leu arg ala ala leu arg ala thr thr ala gly leu arg ala
601/201
cgt gcc cct cca cca tgc tgc tgc
arg pro lys arg ala ser ala cga

```

SEQ ID NO: 33A

FIGURE 33A

104/185

ORF d'après Cole et al. (Nature 393:537-544) et contenant Rv1044

```

1/1                               31/11
taa ttg tgt gca aaa ccg tat cta att gat acg att gcg cac atg gct atc tgg gat cgc
OCH leu cys ala lys pro tyr leu ile asp thr ile ala his met ala ile trp asp arg
61/21                               91/31
ctc gtc gag gtt gcc gcc gag caa cat ggc tac gtc acg act cgc gat gcg cga gac atc
leu val glu val ala ala glu gln his gly tyr val thr thr arg asp ala arg asp ile
121/41                               151/51
ggc gtc gac cct gtg cag ctc cgc ctc cta gcg ggg cgc gga cgt ctt gag cgt gtc gcc
gly val asp pro val gln leu arg leu leu ala gly arg gly arg leu glu arg val gly
181/61                               211/71
cga ggt gtg tac cgg gtg ccc gtg ctg ccg cgt ggt gag cac gac gat ctc gca gcc gca
arg gly val tyr arg val pro val leu pro arg gly glu his asp asp leu ala ala ala
241/81                               271/91
gtg tgg tgg act ttg ggg cgt gcc gtt atc tgg cat gag tgg gcc ttg gcg ctt cat gcc
val ser trp thr leu gly arg gly val ile ser his glu ser ala leu ala leu his ala
301/101                               331/111
ctc gct gac gtg aac ccg tgg cgc atc cat ctc acc gtc cgg cgc aac aac cat ccg cgt
leu ala asp val asn pro ser arg ile his leu thr val pro arg asn asn his pro arg
361/121                               391/131
gcg gcc ggg gcc gag ctg tac cga gtt cac cgc cgc gac ctc cag gca gcc cac ctc act
ala ala gly gly glu leu tyr arg val his arg arg asp leu gln ala ala his val thr
421/141                               451/151
tgg gtc gac gga ata ccc gtc acg acg gtt gcg cgc acc atc aaa gac tgc gtg aag acg
ser val asp gly ile pro val thr thr val ala arg thr ile lys asp cys val lys thr
481/161                               511/171
ggc acg gat cct tat cag ctt cgg gcc gcg atc gag cga gcc gaa gcc gag gcc acg ctt
gly thr asp pro tyr gln leu arg ala ala ile glu arg ala glu ala glu gly thr leu
541/181                               571/191
tgt cgt ggg tta gca ggt gag cta cgc ggt gcg ctc gat gag acc act gcc gga tta cgc
arg arg gly ser ala ala glu leu arg ala ala leu asp glu thr thr ala gly leu arg
601/201
ggt cgg ccg aag cca gga tgg cgg tga
ala arg pro lys arg ala ser ala GGA

```

SEQ ID N° 33F

FIGURE 33F

```

1/1                               31/11
ATA TAA TTA ATT GGT GAT TTT TTT TTT AAC GGA TGA TTA TTT GAT ATT TTT GTC TT, GGA
ATG GTC TT, ala gly TT, ala tta ser asp att arg pr, gly tta att ser leu pro ala
61/21                               91/31
ATA ATA TTT GAA ATA TTT AAC ATT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTA TT, ala, ser, arg, tta tta, ala, ala, ala, ala, ala, ala, ala, ala, ala, ala
121/41                               151/51
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
181/61                               211/71
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
241/81                               271/91
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
301/101                               331/111
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
361/121                               391/131
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
421/141                               451/151
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
481/161                               511/171
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
541/181                               571/191
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
601/201
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
TAA TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT

```

FIG. 33F



105/185

```

1/1                               31/11
TCC AAC CTG CTG GGC CTG CGC CTT CGA ATC GAC GGC CAG GCC ACC GCT CGC TGC CGG CAA
ser asn leu leu gly leu arg leu arg ile asp gly gln ala thr ala arg cys arg gln
61/21                               91/31
CAA CAC CTG GAA TGG GGA CCT TTT CGG TGT TGC TGG TAA CCG GGA CAA CCG GCA CCA CGC
gln his leu glu trp gly pro phe arg cys cys trp OCH pro gly gln pro ala pro arg
121/41                               151/51
CTC GGT CGA GAC GTA TCG CGG CAG CGT TGG CCC TGT CGT TGC TGA CAA TTA CCG CTG GCC
leu gly arg asp val ser arg gln arg trp pro cys arg cys OPA gln leu pro leu ala
181/61                               211/71
GCC GCA TAT TTG CCG CGC TGC CGC GGG CCG GAT C
ala ala tyr leu pro arg cys arg gly pro asp

```

SEQ ID N° 34B

FIGURE 34B

```

1/1                               31/11
GAT CCA ACC TGC TGG GCC TGC GCC TTC GAA TCG ACG GCC AGG CCA CCG CTC GGT GCC GGC
asp pro thr cys trp ala cys ala phe gln ser thr ala arg pro pro leu ala ala gly
61/21                               91/31
AAC AAC ACC TGG AAT GGG GAC CTT TTC GGT GTT GGT GGT AAC CCG GAC AAC CCG CAC CAC
asn asn thr trp asn gly asp leu phe gly val ala gly asn ala asp asn arg his his
121/41                               151/51
GCC TCG CTC GAG ACG TAT CGC GGC AAC GTT GGC CCG GTC GGT GGT GAC AAT TAC CGC TGG
ala ser val glu thr cys arg gly ser val gly pro val val ala asp asn tyr ala trp
181/61                               211/71
CGG CCG CAT ATT TCG CGC GGT GCC GCG GGT GCG ATT
pro pro his ile cys arg ala ala ile arg arg ile

```

SEQ ID N° 34C

FIGURE 34C

106/185

ORF d'après Cole et al. (Nature 393:537-544) contenant seq34A

```

1/1                               31/11
tag ccg cag gcc cct gcg gct agg cgc gcc cgg tgc cgt tgg ccg cgg cgg caa tgc atg
AMB pro gln gly pro ala ala arg arg gly arg cys arg trp pro arg arg gln ser met
61/21                               91/31
ttg cag cag tta caa cgc caa atg gag tct gag cgc atc gtc gag ttc gat cag ctc gcc
leu gln gln leu gln arg gln met glt ser glu arg ile val glu phe asp gln leu gly
121/41                               151/51
agg gga gac gtt gcg cag cga cgg atc caa cct gct ggg cct gcg cct tgc aat cga cgg
arg gly asp val ala gln arg arg ile gln pro ala gly pro ala pro ser asn arg arg
181/61                               211/71
cca gcc cac cgc tgc ctg ccg gca aca aca ccc gga atg ggg acc ttt tgc gtg ttg ctg
pro gly his arg ser leu pro ala thr thr pro gly met gly thr phe ser val leu leu
241/81                               271/91
gta acc ggg aca acc gcc acc acc cct cgg tgc aga cgt atc gcg aca gcg ttg gcc ctg
val thr gly thr thr gly thr thr pro arg ser arg arg ile ala ala ala leu ala leu
301/101                              331/111
tcg ttg ctg aca att acc gct gcc cgc cgc ata ttt gcc gcc ctg ccg cgg gcc gga tcc
ser leu leu thr ile thr ala gly arg arg ile phe ala ala leu pro arg ala gly ser
361/121                              391/131
agg tgc acc tgc cag atc tca ccg cgc agc atc tac gcc gtt cgc tgc aaa ccg ccg act
arg ser thr cys gln ile ser pro arg ser ile tyr ala val arg cys lys pro pro thr
421/141                              451/151
gcg acc gca gcc cca ctc tct tgg cat gcg tcc aat gct gcg acc tcc tgc gta gac aag
ala thr ala gly pro leu ser trp his ala ser asn ala ala thr ser ser val asp lys
481/161                              511/171
ctc acc ctt gcc ttc atg ccg cag tcc tac cca tgt agt aac aga tag
leu thr leu gly phe met pro gln ser tyr pro cys ser asn arg AMB

```

SEQ ID N° 34F

FIGURE 34F

```

1/1                               31/11
CAG TCT CTC GCG AAG GAG GGA CGC ATG GCA CTC TCT GAT CAT GAG CAG CCG ATG CTT GAC
gln ser val gly lys ala gly arg met pro leu ser asp his ala ala arg met leu asp
61/21                               91/31
TAG ATT CAG AAT CTT ATT CAG ACC GAA GAT TTT AAG TTT CCA TCG APT CTC CTT GAT GCG
ala ile gln ser ala leu tyr ala ala asp pro lys phe ala ser ser val arg gly gly
121/41                               151/51
AGT TTT CCA CCA CCA ATT ACC CCA CCG ACC CTT CAG GAT GCG CCG CTT CTT ATG ATG CCG
gly phe ala asp pro thr ala cys arg arg leu ala gly ala ala leu phe ile ala gly
181/61                               211/71
CTG GCG ATG TTT CTT TTT ATT CCG ACC TTC AAA GAG AAT AAT ATT GGA AAT TTT CTT ATA
leu gly cys met val ser gly val ala phe arg ala thr met ile gly asp phe pro ala
241/81                               271/91
CTT AAT CTT TTT CTT CTT ATT CTT AAT CTT TTT CTT CTT CTT CTT CTT CTT CTT CTT CTT
leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr
301/101                              331/111

```

SEQ ID N° 34A

FIGURE 34A

```

1/1
GAC AGT CTG TCG GCA AGG AGG GAC GCA TGC CAC TCT CCG ATC ATG AGC AGC GGA TGC TTG
asp ser leu ser ala arg arg asp ala cys his ser pro ile met ser ser gly cys leu
61/21
ACC AGA TCG AGA GCG CTC TCT ACG CCG AAG ATC CCA AGT TCG CAT CGA GTG TCC GTG GCG
thr arg ser arg ala leu ser thr pro lys ile pro ser ser his arg val ser val ala
121/41
GGG GCT TCC GCG CAC CGA CCG CGC GGC GGC GCC TGC AGG GCG CGG CGT TGT TCA TCA TCG
gly ala ser ala his arg pro arg gly gly ala cys arg ala arg arg cys ser ser ser
181/61
GTC TGG GGA TGT TGG TTT CCG GCG TGG CGT TCA AAG AGA CCA TGA TCG GAA GTT TCC CGA
val trp gly cys trp phe pro ala trp arg ser lys arg pro opa ser glu val ser arg
241/81
TAC TCA GCG TTT TCG GTT TTG TCG TGA TGA TCG GTG GTG TCG TGT ATG CCA TCA CCG GTC
tyr ser ala phe ser val leu ser opa cys ser val val trp cys met pro ser pro val
301/101
CTC GGT TGT CCG GCA GGA TGG ATC GTG GCG GAT CCG CTG CTG GGG CTT CGC GGC AGC GTC
leu gly cys pro ala gly trp ile val ala asp arg leu leu gly leu arg ala ser val.
361/121
GTA CCA AGG GGG CCG GGG GCT CAT TCA CCA GGC GTA TGG AAG ATC
val pro arg gly pro gly ala his ser pro ala val trp lys ile
391/131

```

FIGURE 35B

[illegible]

108/185

séquence RV2169c prédite par Cole et al. (Nature 393:537-544) et contenant partiellement seq35A

1/1	31/11
atg cca ctc tcc gat cat gag cag cgg atg	ctt gac cag atc gag agc gct ctc tac gcc
Met pro leu ser asp his glu gln arg met	leu asp gln ile glu ser ala leu tyr ala
61/21	91/31
gaa gat ccc aag ttc gca tcc agt gtc cgt	ggc ggg ggc ttc cgc gca ccg acc gcg cgg
glu asp pro lys phe ala ser ser val arg	gly gly gly phe arg ala pro thr ala arg
121/41	151/51
cgg cgc ctg cag ggc gcg gcg ttg ttc atc	atc ggt atg ggg atg ttg gtt tcc ggc gtg
arg aig leu gln gly ala ala leu phe ile	ile gly leu gly met leu val ser gly val
181/61	211/71
gcg ttc aaa gag acc atg atc gga agt ttc	ccg ata ttc agc gtt ttc ggt ttt gtc gtg
ala phe lys glu thr met ile gly ser phe	pro ile leu ser val phe gly phe val val
241/81	271/91
atg ttc ggt ggt gtg gtg tat gcc atc acc	ggt cct cgg ttg tcc ggc agg atg gat cgt
met phe gly gly val val tyr ala ile thr	gly pro arg leu ser gly arg met asp arg
301/101	331/111
ggc gga tcc gct gct ggg gct tcc cgc cag	cgt cgt acc aag ggg gcc ggg ggc tca ttc
gly gly ser ala ala gly ala ser arg gln	arg arg thr lys gly ala gly gly ser phe
361/121	391/131
acc agc cgt atg gaa gat cgg ttc cgg cgc	cgc ttc gac gag taa
thr ser arg met glu asp arg phe arg arg	aig phe asp glu OCH

SEQ ID N° 35D

FIGURE 35D

ORF d'après Cole et al. (Nature 393:537-544) et contenant RV2169c

1/1	31/11
gaa cag tcc ctc ggc aag gag gaa cgc agc	cca ctc tcc gat cat gac cag cgg atg ctt
GPA gln ser val gly lys glu gly arg met	pro leu ser asp his glu gln arg met leu
61/21	91/31
gac cag atc gac agc gct ctc tac gcc gaa	gat ccc ctc tcc gat cat gac cag cgg atg
asp gln ile glu ser ala leu tyr ala ala	asp gln asp phe ala ser ser val arg gly
121/41	151/51
ggc gga tcc gct gct ggg gct tcc cgc cag	cgt cgt acc aag ggg gcc ggg ggc tca ttc
gly gly ser ala ala gly ala ser arg gln	arg arg thr lys gly ala gly gly ser phe
181/61	211/71
atg ttc ggt ggt gtg gtg tat gcc atc acc	ggt cct cgg ttg tcc ggc agg atg gat cgt
met phe gly gly val val tyr ala ile thr	gly pro arg leu ser gly arg met asp arg
241/81	271/91
gaa cag tcc ctc ggc aag gag gaa cgc agc	cca ctc tcc gat cat gac cag cgg atg ctt
GPA gln ser val gly lys glu gly arg met	pro leu ser asp his glu gln arg met leu
301/101	331/111
gac cag atc gac agc gct ctc tac gcc gaa	gat ccc ctc tcc gat cat gac cag cgg atg
asp gln ile glu ser ala leu tyr ala ala	asp gln asp phe ala ser ser val arg gly
361/121	391/131
ggc gga tcc gct gct ggg gct tcc cgc cag	cgt cgt acc aag ggg gcc ggg ggc tca ttc
gly gly ser ala ala gly ala ser arg gln	arg arg thr lys gly ala gly gly ser phe
421/141	451/151
acc agc cgt atg gaa gat cgg ttc cgg cgc	cgc ttc gac gag taa
thr ser arg met glu asp arg phe arg arg	aig phe asp glu OCH

ORF ID N° 35D

FIGURE 35D

109/185

```

1/1                               31/11
GAC CTG GGA CGA AGA CGA CGG CAG CAG CCG CAA TCA GAT CTA CCC GGT CCT GGT CAA CGT
asp leu gly arg arg arg arg gln gln pro gln ser asp leu pro gly pro gly gln arg
61/21                               91/31
CAA TGG ACA CCC GAC TAC GGT GCG CCT GCG CGG CTC GAC AAT GCG CGG TTC CTG TTG CCC
gln trp thr pro asp tyr gly ala pro ala arg leu asp asn ala arg phe leu leu pro
121/41                               151/51
GTG GTC GGA GTG CCA CCC GAC CAG GCC ACC GAC TTC GGC TCC GCT GTT GCA CCA GAA ACG
val val gly val pro pro asp gln ala thr asp phe gly ser ala val ala pro glu thr
181/61                               211/71
ACG GCG CCG GTC TGG ATC ACC ATG CTG TGG CCG CTG GCC GAG CGG CCC CGG TTG GCC CCC
thr ala pro val trp ile thr met leu trp pro leu ala asp arg pro arg leu ala pro
241/81                               271/91
CGC GCA CCC GGT GGC ACC GTT CCC GTC CGG CTG GTC GAC GAG GAC CTG GCA AAC TCG CTG
gly ala pro gly gly thr val pro val arg leu val asp asp asp leu ala asn ser leu
301/101                               331/111
GCC AAC GGC GGC CGG CTG GAC ATC CTC CTG TCG GCG GCC GAG TTC GCG ACC AAC CGG GAA
ala asn gly gly arg leu asp ile leu leu ser ala ala glu phe ala thr asn arg glu
361/121                               391/131
GTC GAC CCC GAC GGC GCC GTC GGC CGA GCG CTG TGC CTG GCC ATC GAC CCA GAT C
val asp pro asp gly ala val gly arg ala leu cys leu ala ile asp pro asp

```

SEQ ID N° 36A

FIGURE 36A

```

1/1                               31/11
ACC TGG GAG GAA GAC GAC GGC ACC ACC CGC AAT CAG ATC TAC CCG GTC CTG GTC AAC GTC
thr trp asp glu asp asp gly ser ser arg asn gln ile tyr pro val leu val asn val
61/21                               91/31
AAT GGA CAC CGC ACT AGG GTG GCG CTG CGC GGC TCG ATA ATG GCG GGT TCC TGT TCC GAG
asn gly his pro thr thr val arg leu arg gly ser thr met arg gly ser ty cys pro
121/41                               151/51
TGG TCG GAG TCG CAG GCG AGT AGG GCA CCG ACT TCG GGT GAG CCG CTG TTC GAG CAG AAA CCA
trp ser ala ty his pro thr arg pro thr thr ser ala pro leu leu his ala lys arg
181/61                               211/71
GAG TCG GAG TCG GCA TCA GCA TCC TGT GAG GCG TAC CCG AGT GCG GCG GAT TCG GCG GCG
arg arg arg ser gly ser pro cys tyr gly ala thr pro thr gly pro gly trp pro pro
241/81                               271/91
GAG CAG CCG GTC GCA CCG TCG TCG TCG GGT TCG GAG AGG AGA AAT TCG CAA ACT GTC TCG
gly his phe val ala pro phe pro ser gly trp ser thr ala thr thr ala thr thr thr
301/101                               331/111
AAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG
arg thr ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala

```

SEQ ID N° 36B

FIGURE 36B

1/1	31/11															
CCT GGG ACG AAG ACG ACG GCA GCA GCC GCA	ATC AGA TCT ACC CGG TCC TGG TCA ACG TCA															
pro gly thr lys thr thr ala ala ala ala	ile arg ser thr arg ser trp ser thr ser															
61/21	91/31															
ATG GAC ACC CGA CTA CGG TGC GCC TGC GCG	GCT CGA CAA TGC GCG GTT CCT GTT GCC CGT															
met asp thr arg leu arg cys ala cys ala	ala arg gln cys ala val pro val ala arg															
121/41	151/51															
GGT CGG AGT GCC ACC CGA CCA GGC CAC CGA	CTT CGG CTC CCC TGT TCC ACC AGA AAC GAC															
gly arg ser ala thr arg pro gly his arg	leu arg leu arg cys cys thr arg asn asp															
181/61	211/71															
CCC GCC GGT CTG GAT CAC CAT GCT GTG GCC	GCT GGC CGA CCG GCC CCG GTT GGC CCC CGG															
gly ala gly leu asp his his ala val ala	ala gly arg pro ala pro val gly pro arg															
241/81	271/91															
CCC ACC CGS TGS CAC CGT TCC CST GCG GCT	GCT CGA CGA CGA CCT GGC AAA CTC GCT GGC															
gly thr arg trp his arg ser arg pro ala	gly arg arg arg pro gly lys leu ala gly															
301/101	331/111															
CAA CGG CGS CGS GCT GSA CAT CST CCT GTC	GCC GGC CGA GTT CGC CAC CAA CGG GGA AGT															
gin arg arg pro ala gly his pro pro val	gly gly arg val arg his gin pro gly ser															
361/121	391/131															
CGA CCC CGA CGS CGC CST CGS CCG AGC GCT	GTG CCT GGC CAT CGA CCC AGA TC															
arg pro arg arg arg arg arg pro ser ala	val pro gly his arg pro arg															

FIGURE 36C

361/121  
GTC GCC CCC GAA CTA GAC CGC GGG CAA GAG GCC GGC TTT ACC CTC TCG GCC CCG CTG CGC  
val ala pro glu leu asp arg gly gln glu ala gly phe thr leu ser ala pro leu arg  
421/141  
TCG CTG ACC AGG CCG TCG TTG GCC GTC AAC CAG CCC GGG ATC TAC CCG GTC CTG GTC AAC  
ser leu thr arg pro ser leu ala val asn gln pro gly ile tyr pro val leu val asn  
481/161  
GTC AAT GGG ACA CCC GAC TAC GGT GCG CCT GCC CGG CTC GAC AAT GCG CGG TTC CTG TTG  
val asn gly thr pro asp tyr gly ala pro ala arg leu asp asn ala arg phe leu leu  
541/181  
CCC GTG GTC GGA GTG CCA CCC GAC CAG GCC ACC GAG TTC GGC TCC GCT GTT GCA CCA GAA  
pro val val gly val pro pro asp gln ala thr asp phe gly ser ala val ala pro glu  
601/201  
ACG ACG GCG CCG CTC TGG ATC ACC ATG CTC TGG CCG CTG GCC GAC CCG CCC CCG TTG GCC  
thr thr ala pro val trp ile thr met leu trp pro leu ala asp arg pro arg leu ala  
661/221  
CCG GGG GCA CCC GGT GGC ACC GTT CCC GTA CCG CTG GTC GAC SAC GAC CTC GCA AAC TCG  
pro gly ala pro gly gly thr val pro val arg leu val asp asp asp leu ala asn ser  
721/241  
CTG GCG AAC GGG GGC CCG CTC GAC ATC CTC CTG TCG CCG GCC GAG TTC GCC ACC AAC CCG  
leu ala asa gly gly arg leu asp ile leu leu ser ala ala glu phe ala thr asn arg  
781/261  
GAA GTG GAC CCC GAC GGC GCG GTC GCG CSA GCG CTG TGC CTG GCG ATC GAC CCA GAT CTA  
glu val asp pro asp gly ala val gly arg ala leu cys leu ala ile asp pro asp leu  
841/281  
CTC ATC ACC GTC AAT GCG ATG ACC GGC GGC TAG GTC GTG TCC GAC TCG CCC GAC GCG GCC  
leu ile thr val asn ala met thr gly gly tyr val val ser asp ser pro asp gly ala  
901/301  
GCT CAA CTA CCG GGC ACC CCG ACC CAC CCG GCG ACC GCG CAG GCG GCC SCA TCC ASC TGG  
ala gln leu pro gly thr pro thr his pro gly thr gly gln ala ala ala ser ser trp  
961/321  
CTG GAT CCA TTG CCG ACC CTA GTC CAC CGA ACA TGC GTG ACC CCG CTC CCT TTT GCG CAA  
leu asp arg leu arg thr leu val his arg thr cys val thr pro leu pro phe ala gln  
1021/341  
GTC GAC CTC GAT GGT TTG CAG CCG GTT AAA GAT CCG ACC CTC AAT GAT ATA GCA ATT ATT  
ala asp leu asp ala leu gln arg val asn asp pro arg leu ser ala ile ala thr ile  
1081/361  
AGC GCG GCG GAG ATC CTC TAC CCG ATG CTC GAT GGT AGT TAT AAT ATT ATT ALA A ...  
ser pro ala asp ile val asp arg ile leu arg val ser thr thr thr thr thr thr  
1141/381  
CTG CTC GAT GGT GCG TTA AAT GAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT  
leu pro asp thr pro thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
1201/401  
...  
1261/421  
...  
1321/441  
...  
1381/461  
...  
1441/481  
...  
1501/501

1441/481  
GAC GCC TTG GGC GCA ATG CTG TGG CGC AGC TTG GAG CCG AAT GCC GCG CCC CGT ACC CAA  
asp ala leu gly ala met leu trp arg ser leu glu pro asn ala ala pro arg thr gln  
1501/501  
ATC CTG GTG CCG CCG GCG TCG TGG AGC CTG GCC AGC GAC GAC GCG CAG GTC ATC CTG ACC  
ile leu val pro pro ala ser trp ser leu ala ser asp asp ala gln val ile leu thr  
1561/521  
GGG CTG GCC ACC GCC ATC CGG TCT GGC CTG GCC GTG CCG CGA CCA CTA CCG GCG GTG ATC  
ala leu ala thr ala ile arg ser gly leu ala val pro arg pro leu pro ala val ile  
1621/541  
GCT GAC GCC GCG GCC CGC ACC GAG CCA CCG GAA CCG CCG GGC GCT TAC AGC GCC GCT CGC  
ala asp ala ala ala arg thr glu pro pro glu pro pro gly ala tyr ser ala ala arg  
1681/561  
GGC CGG TTC AAT GAC GAC ATC ACC ACG CAG ATC GGC GGG CAG GTT GCC CCG CTA TGG AAG  
gly arg phe asn asp asp ile thr thr gln ile gly gly gln val ala arg leu trp lys  
1741/581  
CTG ACC TCG GCG TTS ACC ATC GAT GAC CGC ACC GGG CTG ACC GGC GTG CAG TAC ACC GCA  
leu thr ser ala leu thr ile asp asp arg thr gly leu thr gly val gln tyr thr ala  
1801/601  
GCA CTA CGC GAG GAC ATG TTG GCG GCG CTG AGC CAA TCG CTA CCA CCC GAT ACC CGC AAC  
pro leu arg glu asp met leu arg ala leu ser gln ser leu pro pro asp thr arg asn  
1861/621  
GGG CTG GGC CAG CAG CCG CTG GGC GTC GTT CGA AAG ACG ATC GAC SAT CTT TTC GGC GCG  
gly leu ala gln gln arg leu ala val val gly lys thr ile asp asp leu phe gly ala  
1921/641  
GTG ACC ATC GTC AAC CCG GGC GGC TCG TAC ACT CTG GGC ACC GAG CAC AGT CCG CTG CCG  
val thr ile val asn pro gly gly ser tyr thr leu ala thr glu his ser pro leu pro  
1981/661  
TTG GCG CTG GAT AAT GGC CTG GGC GTC CCA ATC CCG GTG CCG CTA CAG GTC GAT GCT CGC  
leu ala leu his asn gly leu ala val pro ile arg val arg leu gln val asp ala pro  
2041/681  
CCG GCG ATG ACG CTG GGC GAG GTC GGT CAG ATC CAG GTA CCG CCG GCG TAC CTG CCG CTA  
pro gly met thr val ala asp val gly gln ile ile leu pro pro gly tyr leu pro leu  
2101/701  
GGA GTA CCA ATC CAG GTG AA TTG ACA CAG CCG GTT GCG GTC GAG CTG TCG CTG CCG ACC  
arg val pro ile glu val asn phe thr gln arg val ala val asp val ser leu arg thr  
2161/721  
CTT GAC GAT GCG GCG GCG GCG GAA GCG GTC GAT CTG TCG GCG GAG CCG AA GCG TAC GCG GCG  
pro asp gly val ala leu gly gln pro val arg leu ser val his leu pro ala tyr gly  
2221/741  
AAT GTT TTT TTT TTT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
lys val leu leu ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala  
2281/761  
GCT TTT TTT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2341/781  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2401/801  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2461/821  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2521/841  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2581/861  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2641/881  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2701/901  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2761/921  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2821/941  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2881/961  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
2941/981  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
3001/1001  
GCT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT AAT  
arg leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr



113/185

ORF d'après Cole et al., 1998 (Nature 393 537-544) et contenant Rv 3909.

1/1	31/11																		
TGA CTC AGC ACC GGG TCA GCA CAA CGG TCC	CGG GCC GGG GCC GTG ACC GCA CTG CAA CTC																		
OPA leu ser thr gly ser ala gln arg ser	arg ala gly ala val thr ala leu gln leu																		
61/21	91/31																		
GGC TGG GCC GCT TTG CCG CGC GTC ACC TCA	GGC ATC GGC GTC GTG GCC GGC CTC GGG ATG																		
gly trp ala ala leu ala arg val thr ser	ala ile gly val val ala gly leu gly met																		
121/41	151/51																		
GGG CTC ACG GTA CCG TCG GCG GCA CCG CAC	GGG CTC GCA GGC GAG CCC AGC CCG ACG CCT																		
ala leu thr val pro ser ala ala pro his	ala leu ala gly glu pro ser pro thr pro																		
181/61	211/71																		
TTT GTC CAG GTC CGC ATC GAT CAG GTG ACC	CGG GAC GTG GTG ACC ACT TCC AGC GAA CCC																		
phe val gln val arg ile asp gln val thr	pro asp val val thr thr ser ser glu pro																		
241/81	271/91																		
CAT GTC ACC GTC AGC GGA ACG GTG ACC AAT	ACC GGT GAC GGC CCA GTC CGC GAT GTG ATG																		
his val thr val ser gly thr val thr asn	thr gly asp arg pro val arg asp val met																		
301/101	331/111																		
CTC CGG CTT GAG CAC GGC GGC GCG GTC ACC	TCG TCA ACG GCG TTA CGC ACC TCG CTC GAC																		
val arg leu glu his ala ala ala val thr	ser ser thr ala leu arg thr ser leu asp																		
361/121	391/131																		
GGC GGC ACC GAC CAG TAC CAG CCG GGC GCG	GAC TTA CTC AAG GTC GCG CCC GAA CTA GAC																		
gly gly thr asp gln tyr gln pro ala ala	asp phe leu thr val ala pro glu leu asp																		
421/141	451/151																		
CGC GGG CAA GAG GCC GGC TTT ACC CTC TCG	GGC CCG CTC CGC TCG CTC ACC AGG CCG TCG																		
arg gly gln glu ala gly phe thr leu ser	ala pro leu arg ser leu thr arg pro ser																		
481/161	511/171																		
TTG GGC GTC AAC CAG CCC GGG ATC TAC CCG	GTG CTG GTC AAG CTC AAT GGG ACA CCC GAC																		
leu ala val asn gln pro gly ile tyr pro	val leu val asn val asn gly thr pro asp																		
541/181	571/181																		
TAC GGT CCG GGT GCG CCG CTC GAC AAT GGC	CTG TTT CTC TCG GAG CTC CTC GGA GTG CCA																		
tyr gly ala pro ala asn leu asp asn ala	arg asp pro leu thr val val gly val pro																		
631/201	661/211																		
CGC GAC CAG GAT ACG GAT TTC GGC TCG GTT	CTT TTA CCA GAA AAT AAT GAT CCG CTC TCA																		
pro asp gln ala thr asp phe gly ser ala	val ala pro ile thr thr ala pro val tyr																		
661/221	691/231																		
ATC AAG ATA GAG TCG GAT CTC GAT GAT GAT	CTG GAT GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr met asp thr pro leu ala asp his	ile thr met asp thr pro leu ala asp his																		
711/241	741/251																		
AAT TTT CCG GGT GAT GAT GAT GAT GAT GAT	CTT TTA AAT GAT GAT GAT GAT GAT GAT GAT																		
asn val pro asn asp asp asp asp asp asp	thr thr asn asp asp asp asp asp asp asp																		
741/261	771/271																		
GAT GAT AAT GAT GAT GAT GAT GAT GAT GAT	CTT AAT AAT GAT GAT GAT GAT GAT GAT GAT																		
leu asp ile leu leu asp ala ala ile thr	thr thr thr thr thr thr thr thr thr thr																		
841/281	871/291																		
GAT TTT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
871/301	901/311																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
901/321	931/331																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
931/341	961/351																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
961/361	991/371																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
991/381	1021/391																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1021/401	1051/411																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1051/421	1081/431																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1081/441	1111/451																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1111/461	1141/471																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1141/481	1171/491																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1171/501	1201/511																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1201/521	1231/531																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1231/541	1261/551																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1261/561	1291/571																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1291/581	1321/591																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1321/601	1351/611																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1351/621	1381/631																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1381/641	1411/651																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT GAT GAT GAT GAT GAT GAT GAT																		
ile thr thr thr thr thr thr thr thr thr thr	thr thr thr thr thr thr thr thr thr thr																		
1411/661	1441/671																		
GAT AAT GAT GAT GAT GAT GAT GAT GAT GAT	CTT TTA GAT																		

114/185

[illegible]

115/185

2041/681  
 CTC GCC GTG CCA ATC CGG GTC CGG CTA CAG GTC GAT GCT CCG CCC GGG ATG ACG GTG GCC  
 leu ala val pro ile arg val arg leu gln val asp ala pro pro gly met thr val ala  
 2101/701  
 GAT GTC GGT CAG ATC GAG CTA CCG CCC GGG TAC CTG CCG CTA CGA GTA CCA ATC GAG GTG  
 asp val gly gln ile glu leu pro pro gly tyr leu pro leu arg val pro ile glu val  
 2161/721  
 AAC TTC ACA CAG CGG GTT GCC GTC GAC GTG TCG CTG CCG ACC CCC GAC GGC GTC GCG CTG  
 asn phe thr gln arg val ala val asp val ser leu arg thr pro asp gly val ala leu  
 2221/741  
 GGT GAA CCG GTG CCG TTG TCG GTG CAC TCC AAC GCC TAC GGC AAG GTG TTG TTC GCG ATC  
 gly glu pro val arg leu ser val his ser asn ala tyr gly lys val leu phe ala ile  
 2281/761  
 ACG CTA TCC GCT GCG GCC GTG CTC CTA ACG CTG GCG GCG CCG GCG CTT TGG CAC CCG TTC  
 thr leu ser ala ala ala val leu val thr leu ala gly arg arg leu trp his arg phe  
 2341/781  
 CGT GGC CAG CCG GAT CCG GCC GAC CTG GAT CCG CCG GAC CTG CCG ACC GGC AAA CAC GCC  
 arg gly gln pro asp arg ala asp leu asp arg pro asp leu pro thr gly lys his ala  
 2401/801  
 CCG CAG CCG CCG GCC GTA GCC AGT CCG GAT GAC GAA AAG CAC CCG GTA TGA  
 pro gln arg arg ala val ala ser arg asp asp ala lys his arg val opa

SEQ ID 36F (suite 2)

FIGURE 36F (suite 2)

1/1  
 ATC CCG GCG TTS GCG TCG GAT CCG AAC ATC CTC GGA GTC AAG GAC GCG AAA GCG GAC CTC  
 ile arg ala leu ala ser his pro asn ile val gly val lys asp ala lys ala asp leu  
 61/21  
 CAC AGC GCG GCC CAA ATC ATG GCG GAC ACC GGA CTC GCG TAC TAT TCC GCG GAC GAC GCG  
 his ser gly ala gln ile met ala asp thr gly leu ala tyr tyr ser gly asp asp ala  
 121/41  
 CTC AAC CTC TAT TAT CTC GCG AAG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG  
 leu asn leu pro thr leu ala met gly ala thr gly phe ile ser val ile ala his leu  
 181/61  
 GCA GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG GCG  
 ala ala val thr leu ala his leu leu leu leu leu leu leu leu leu leu leu leu leu  
 241/81  
 TCG AAC ATC  
 thr lys ile

SEQ ID N 37A

116/185

```

1/1                               31/11
GAT CCG CGC GTT GGC GTC GCA TCC GAA CAT CGT CGG AGT CAA GGA CGC CAA AGC CGA CCT
asp pro arg val gly val ala ser glu his arg arg ser gln gly arg gln ser arg pro
61/21                               91/31
GCA CAG CGG CGC CCA AAT CAT GGC CGA CAC CGG ACT GGC CTA CTA TTC CGG CGA CGA CGC
ala gln arg arg pro asn his gly arg his arg thr gly leu leu phe arg arg arg arg
121/41                               151/51
GCT CAA CCT GCC CTG GCT GGC CAT GGG CGC CAC GGG GTT CAT CAG CGT GAT TGC CCA CCT
ala gln pro ala leu ala gly his gly arg his gly leu his gln arg asp cys pro pro
181/61                               211/71
GGC AGC CGG GCA GCT TCG AGA GTT GTT GTC CGC CTT CGG TTC TGG GGA TAT CGC CAC CGC
gly ser arg ala ala ser arg val val val arg leu arg phe trp gly tyr arg his arg
241/81
CCG CAA GAT C
pro gln asp

```

SEQ ID N° 37B

FIGURE 37B

```

1/1                               31/11
TCC GCG CGT TGG CGT CGC ATC CGA ACA TCG TCG GAG TCA AGG ACG CCA AAG CCG ACC TGC
ser ala arg trp arg arg ile arg thr ser ser glu ser arg thr pro lys pro thr cys
61/21                               91/31
ACA GCG GCG CGC AAA TCA TCG CGC ACA CGC GAG TGG CCG ACT ATT CCG GCG ACG ACG CGC
thr ala ala pro lys ser trp pro thr pro arg trp pro thr ile pro ala thr thr arg
121/41                               151/51
TCA ACC TCG CCG GCG TCG CCA TGG GCG CGA CGC GTT TCA TCA GCG TCA TCG CCG ACC TGG
ser thr cys pro gly trp pro trp ala pro arg ala ser ser ala asp leu pro thr thr
181/61                               211/71
CAG CCG GCG ACC TTC GAG ACT TGT TGT CCG CCG TCG GTT CTG GAG ATA TCG CCA TCG CCG
thr pro gly ser pro ala cys cys pro pro leu val leu gly ile ser pro thr pro
241/81
ACA ACA T
acc ala

```

SEQ ID N° 37C

FIGURE 37C

117/185

Séquence codante Rv2753c prédite par Cole et al., 1998 (Nature 393 537-544)  
contenant Seq 37A

```

1/1                               31/11
GTG ACC ACC GTC GGA TTC GAC GTC GCA GCG CGC CTA GGA ACC CTG CTG ACC GCG ATG GTG
val thr thr val gly phe asp val ala ala arg leu gly thr leu leu thr ala met val
61/21                               91/31
ACA CCG TTT AGC GGC GAT GGC TCC CTG GAC ACC GCG ACC GCG GCG CTG GCC AAC CAC
thr pro phe ser gly asp gly ser leu asp thr ala thr ala ala arg leu ala asn his
121/41                               151/51
CTG GTC GAT CAG GGG TGC GAC GGT CTG GTG GTC TCG GCG ACC ACC GCG GAG TCG CCG ACC
leu val asp gln gly cys asp gly leu val val ser gly thr thr gly glu ser pro thr
181/61                               211/71
ACC ACC GAC GGG GAG AAA ATC GAG CTG CTG CCG GCG GTC TTG GAA GCG GTG GGG GAC CCG
thr thr asp gly glu lys ile glu leu leu arg ala val leu glu ala val gly asp arg
241/81                               271/91
GCG CGT GTT ATC GCG GGT GCG GCG ACC TAT GAC ACC GCG CAG AGC ATC CCG CTG GCG AAG
ala arg val ile ala gly ala gly thr tyr asp thr ala his ser ile arg leu ala lys
301/101                               331/111
GGT TGT GCG GCG GAG GGT GCG GAG GCG CTG CTG GCG GTC ATC GCG TAC TAT TCG AAG CCG
ala cys ala ala glu gly ala his gly leu leu val val thr pro tyr tyr ser lys pro
361/121                               391/131
CCG CAG CCG GCG CTG CAA GCG CAT TTC ACC GCG GTC GGT GAG GCG ACC GAG CTG CCG ATG
pro gln arg gly leu gln ala his phe thr ala val ala asp ala thr glu leu pro met
421/141                               451/151
CTG CTC TAT GAC ATC CCG GGG CCG TCG GCG GTG CCG ATT GAG GCG GAG ACC ATC CCG GCG
leu leu tyr asp ile pro gly arg ser ala val pro ile glu pro asp thr ile arg ala
481/161                               511/171
TTG GCG TCG CAT CCG AAC ATC GTC GGA GTC AAG GAG GAT AAA GCG GAC CTG CAG AGC GCG
leu ala ser his pro asn ile val gly val lys asp ala lys ala asp leu his ser gly
541/181                               571/191
GCG CAA ATC ATG GCG GAC ACC GGA CTG GCG TAT TAT TTT GAT GAT GAG GCG CTG AAC CTG
ala gln ile met ala asp thr gly leu ala tyr tyr leu gly asp asp ala leu asn leu
601/201                               631/211
CCG TGG CTG GCG ATG GCG GAT ATC GAT TTG AAT AAT GAT GAT GAT GAT GAT GAT GAT GAT
pro trp leu ala met gly ala thr gly phe ile ser thr ala ala his leu ala ala gly
661/221                               691/231
CAG GTT CCA GAG TTG TTG TCG GCG TTG GGT TAT CGA GAT ATC GGT ACC GAT CGA AA. ATC
gln leu arg gln leu leu ser ala pro gly ser ala thr ala thr ala thr ala thr ala
721/241                               751/251
AAT ATT GAT GTT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
asn ile ala val ala pro leu cys ala ala ser thr thr thr thr thr thr thr thr thr
781/261                               811/271
AAT GGT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
asn gln ala val ala pro leu cys ala ala ser thr thr thr thr thr thr thr thr thr
841/281                               871/291
GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
asn gln ala val ala pro leu cys ala ala ser thr thr thr thr thr thr thr thr thr
901/301                               931/311
GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
asn gln ala val ala pro leu cys ala ala ser thr thr thr thr thr thr thr thr thr

```

ORF d'après Cole et al., 1998 (Nature 393 537-544) contenant Hv2753c

[illegible]

119/185

1021/341  
 GAC GGG GAG AAA ATC GAG CTG CTG CGG GCC GTC TTG GAA GCG GTG GGG GAC CGG GCC CGT  
 asp gly glu lys ile glu leu leu arg ala val leu glu ala val gly asp arg ala arg  
 1081/361  
 GTT ATC GCC GGT GCC GGC ACC TAT GAC ACC GCG CAC AGC ATC CGG CTG GCC AAG GCT TGT  
 val ile ala gly ala gly thr tyr asp thr ala his ser ile arg leu ala lys ala cys  
 1141/381  
 GCG GCC GAG GGT GCG CAC GGG CTG CTG GTG GTC ACG CCC TAC TAT TCC AAG CCS CCG CAG  
 ala ala glu gly ala his gly leu leu val val thr pro tyr tyr ser lys pro pro gln  
 1201/401  
 CGG GGG CTG CAA GCC CAT TTC ACC GCC GTC GCC GAC GCG ACC GAG CTG CCS ATG CTG CTC  
 arg gly leu gln ala his phe thr ala val ala asp ala thr glu leu pro met leu leu  
 1261/421  
 TAT GAC ATC CCG GGG CCG TCG GCG GTG CCC ATC GAG CCC GAC ACG ATC CCG CCG TTG GCG  
 tyr asp ile pro gly arg ser ala val pro ile glu pro asp thr ile arg ala leu ala  
 1321/441  
 TCG CAT CCG AAC ATC CTC GGA GTC AAG GAC GCC AAA GCG GAC CTG CAC AGC GCG GCC CAA  
 ser his pro asn ile val gly val lys asp ala lys ala asp leu his ser gly ala gln  
 1381/461  
 ATC ATG GCC GAC ACC GGA CTG GCG TAC TAT TCC GCG GAC GAC GCG CTC AAC CTG CCG TGG  
 ile met ala asp thr gly leu ala tyr tyr ser gly asp asp ala leu asn leu pro trp  
 1441/481  
 CTG SCC ATG GCG GCC ACG GCG TTC ATC AGC GTG ATT GCC CAC CTG GCA GCC GGG CAG CTT  
 leu ala met gly ala thr gly phe ile ser val ile ala his leu ala ala gly gln leu  
 1501/501  
 CGA GAG TTG TTG TCC GCG TTC GGT TCT GCG GAT ATC GCG ACC GCG CCG AAG ATC AAC ATT  
 arg glu leu leu ser ala phe gly ser gly asp ile ala thr ala arg lys ile asn ile  
 1561/521  
 GCG GTC GCG CCG CTG TCG AAC GCG ATG ACC CCG CTG GGT GCG GTG ACG TTG TCC AAG GCG  
 ala val ala pro leu tyr asn ala met ser arg leu gly gly val thr leu ser lys ala  
 1621/541  
 GCG TTG CCG CTG CAG GCC ATC GAC GTC GGT GAT GCG CCG CTG CCG CAG GTG GGT GCG ACA  
 gly leu arg leu gln gly ile asp val gly asp pro arg leu pro gln val ala ala thr  
 1681/561  
 CCG GAG CAG ATC GAT GCG TTG GCG GCG GAT ATG CCG CCG GAG TCG GTG CTT CCG TGA  
 pro glu gln ile asp ala leu ala ala asp met arg ala ala ser val leu arg cpa

SEQ ID NO 37F (suite 1)

FIGURE 37F (suite 1)





121/185

```

1/1
AGC GGT GAA CTG GTG GGC CCG GAT GGT TCA
ser gly glu leu val gly pro asp gly ser
61/21
GAG ACG ACG GAT GGA AGG AGA TGC TGG CGC
glu thr thr asp gly arg arg cys trp arg
121/41
GTC GTC GTC TAC GGA GGT GAA GGC GGC GGA TTC GGC GGA ATC TGA CGC CGG AGC CGA CCA
val val val tyr gly gly glu gly gly gly phe gly gly ile OPA arg arg ser arg pro
181/61
GAC TGG CCC GCA GGT GAA GGC GGC GGA TTC GGC GGA ATC TGA CGC CGG AGA GCT CGG CGA
asp trp pro ala gly glu gly gly gly phe gly gly ile OPA arg arg arg ala arg arg
241/81
GGA CGC GTG CCC AGA ACA GGC CCG CGT CGA GGC GGC CGC GTC GCG GTT GCG GCG AGG CTG
gly arg val pro arg thr gly pro arg arg ala ala pro val ala val ala ala arg leu
301/101
GCT TGT TGG CAT TGC GGC GAC GGT GGT CGC GTT GGT GGC TGG CCG TGG CGC AGC GGG TTA
ala cys trp his cys gly asp ala ala arg val gly arg trp pro trp arg ser gly leu
361/121
TTT TGC GTT GCG CTC ACA CCA GGA AAG CCA ATC AAT GGC GGC CGA GSA CCG TGC GGC CAT
phe cys val ala leu thr pro gly lys pro ile asn arg ala arg gly pro cys gly his
421/141
TGA GGC CGC TAA GSA TTG CGT TGC GGC CAC GCA GGC AGC CGA TGC TGG GGC GAT GTC GGC
OPA gly arg OCH gly leu arg cys gly his ala gly thr arg cys trp gly asp val gly
481/161
TAG CAT GCA GAA GAT C
AAB his ala glu asp

```

SEQ ID NO 380

FIGURE 380

122/185

Séquence Rv0175 prédite par Cole et al., 1998 (Nature 393 537-544) et contenant seq38A

```

1/1                               31/11
GTG AAG GCG GCG GAT TCG GCG GAA TCT GAC GCC GGA GCC GAC CAG ACT GGC CCG CAG GTG
val lys ala ala asp ser ala glu ser asp ala gly ala asp gln thr gly pro gln val
61/21                               91/31
AAG GCG GCG GAT TCG GCG GAA TCT GAC GCC GGA GAG CTC GGC GAG GAC GCG TGC CCA GAA
lys ala ala asp ser ala glu ser asp ala gly glu leu gly glu asp ala cys pro glu
121/41                               151/51
CAG GCC CTC GTC GAG CGG CGC CCG TCG CGG TTG GCG CGA GGC TGG CTT GTT GGC ATT GCG
gln ala leu val glu arg arg pro ser arg leu arg arg gly trp leu val gly ile ala
181/61                               211/71
GCG ACG CTG CTC GCG TTG GCC GGT GGC CTT GGC GCA GCG GGT TAT TTT GCG TTG GCG TCA
ala thr leu leu ala leu ala gly gly leu gly ala ala gly tyr phe ala leu arg ser
241/81                               271/91
CAC CAG GAA AGC CAA TCA ATC GCG CGC GAG SAC GTT GCG GCG ATT GAG GCG GCT AAG GAT
his gln glu ser gln ser ile ala arg glu asp leu ala ala ile glu ala ala lys asp
301/101                               331/111
TGC GTT GCG GCG ACC CAG GCA GCG GAT GCT GCG GCG ATG TCG GCT AGC ATG CAG AAG ATC
cys val ala ala thr gln ala pro asp ala gly ala met ser ala ser met gln lys ile
361/121                               391/131
ATC GAG TGT GCG ACC GGT GAT TTC GGT GCG CAG GCG TCG TTG TAC ACC AGC ATG CTC CTC
ile glu cys gly thr gly asp phe gly ala gln ala ser leu tyr thr ser met leu val
421/141                               451/151
GAG GCG TAT CAA GCG GCC AGC GTC CAC GTG CAA GTG ACC GAT ATG CCG GCG GCG GTC GAG
glu ala tyr gln ala ala ser val his val gln val thr asp met arg ala ala val glu
481/161                               511/171
CGC AAC AAC AAT GAC GGG TCG GTG GAT GTT CTG GTG GCG CTC CGG GTC AAG GTG TCC AAC
arg asn asn asp gly ser val asp val leu val ala leu arg val lys val ser asn
541/181                               571/191
ACC GAC TCG GAT GCG CAT GAA GTC GAT TAC GTT GTT GCG GTC CGG ATG GCA CTC GAT GAG
thr asp ser asp ala his glu val gly tyr arg leu arg val arg met ala leu asp ala
601/201                               631/211
GGC GCG TAT AAG ATC GCG AAA CTC GAG CAG GTT AAG AAA TCA
gly arg tyr lys ile ala lys leu asp ala val thr lys gln

```

SEQ ID N 381

FIGURE 381

123/185

ORF d'après Cole et al., 1998 (Nature 393 537-544) Contenant Rv0175

```

1/1                                31/11
TGA ACT GGT GGG GCC GGA TGG TGT CAA GTA CGC CGT CGC AAA CTC GAG CAC AAC AGG AGA
OPA thr gly gly ala gly trp cys gln val arg arg arg lys leu glu his asn arg arg
61/21                                91/31
CGA CGG ATG GAA GGA GAT GCT GGC GCC GGC CAG CTG AAC CCT GCC GAT GCG AAT AAG TCG
arg arg met glu gly asp ala gly ala gly gln leu asn pro ala asp ala asn lys ser
121/41                                151/51
TCG TCT ACG GAG GTG AAG GCG GCG GAT TCG GCG GAA TCT GAC GCC GGA GCC GAC CAG ACT
ser ser thr glu val lys ala ala asp ser ala glu ser asp ala gly ala asp gln thr
181/61                                211/71
GGC CCG CAG GTG AAG GCG GCG GAT TCG GCG GAA TCT GAC GCC GGA GAG CTC GCC GAG GAC
gly pro gln val lys ala ala asp ser ala glu ser asp ala gly glu leu gly glu asp
241/81                                271/91
GCG TGC CCA GAA CAG GCC CTC CTC GAG CCG GCG GCG TCG GCG TTG CCG CGA GCG TCG CTT
ala cys pro glu gln ala leu val glu arg arg pro ser arg leu arg arg gly trp leu
301/101                                331/111
GTT GGC ATT GCG GCG ACG CTC CTC GCG TTG GCG GGT GCG CTT GCG CCA GCG GGT TAT TTT
val gly ile ala ala thr leu leu ala leu ala gly gly leu gly ala ala gly tyr phe
361/121                                391/131
GCG TTG CCG TCA CAC CAG GAA ACC CAA TCA ATC GCG GCG GAG GAC CTT GCG GCC ATT GAG
ala leu arg ser his gln glu ser gln ser ile ala arg glu asp leu ala ala ile glu
421/141                                451/151
GCC GCT AAG GAT TGC GTT GCG GCC ACC CAG GCA CCC GAT GGT GGG GCG ATC TCG GCT AGC
ala ala lys asp cys val ala ala thr gln ala pro asp ala gly ala met ser ala ser
481/161                                511/171
ATG CAG AAG ATC ATC GAG TGT GGC ACC GGT GAT TTC GGT GCG CAG GCG TCG TTG TAC ACC
met gln lys ile ile glu cys gly thr gly asp phe gly ala gln ala ser leu tyr thr
541/181                                571/191
ACC ATG CTC GTC GAG GCG TAT CAA GCG GCG ACC GTC CAT GTC CAA GTC ACC GAT AAG CCG
ser met leu val glu ala tyr gln ala ala ser val his val gln val thr asp met arg
601/201                                631/211
TCG GCG GTT GAG CCG AAG AAC AAT GAG GCG TCG GTC GAT GGT CTC GTC GCG CTC GCG GTT
ala ala val glu arg asn asn asp asp gly ser val asp val leu val ala leu arg val
661/221                                691/231
AAG GTC TCG AAG ACC GAC TCG GAT GCG CAT CAA CTC GCG TAT GGT TTT GCG TTT GCG AAG
lys val ser asn thr asp ser asn ala his gln val gly thr arg leu thr val ser met
721/241                                751/251
GCA CTG GAT GAG GAT TTT TAT AAG AAT GGT AAA CTC GAT TAT CTC AAT AAG TCA
ala leu arg ala gly arg tyr leu ala gln ala gln val thr thr leu arg val thr lys thr

```

SEQ ID NO 39F

PIVIER 39F

124/185

1/1 31/11  
ACA CCT CCC CCC CCG CCG CCG CTG CCG CCG GTT CCC TTT CCC AAG GAA TGT CCG GCG CCG  
thr pro pro pro pro pro pro leu pro pro val pro phe pro lys glu cys pro ala pro  
61/21 91/31  
GGC GTG ATG CAA GGC TGC CTT GAG AGC ACC AGC GGC TTG ATC ATG GGC ATC GAC AGC AAG  
gly val met gln gly cys leu glu ser thr ser gly leu ile met gly ile asp ser lys  
121/41 151/51  
ACC GCA CTG GTC GCC GAG CGC ATC ACC GGT GCC GTC GAG GAG ATC  
thr ala leu val ala glu arg ile thr gly ala val glu gluile

SEQ ID N° 39A

FIGURE 39A

1/1 31/11  
CAC CTC CCC CCC CGC CGC TGC CGC CGG TTC CCG TTC CCA AGG AAT GTC CGG CGG CGG  
his leu pro pro arg arg arg cys arg arg phe pro phe pro arg asn val arg arg arg  
61/21 91/31  
GCG TGA TGC AAG GCT GGC TTG AGA GCA CCA GCG GGT TGA TCA TGG GCA TCG ACA GCA AGA  
ala OPA cys lys ala ala leu arg ala pro ala ala OPA ser trp ala ser thr ala arg  
121/41 151/51  
CCG CAC TGG TCG CGG AGC GCA TCA CCG GTG CCG TCG AGG AGA TC  
pro his trp ser pro ser ala ser pro val pro ser arg arg

SEQ ID N° 39B

FIGURE 39B

125/185

Séquence codante Rv3006 prédite par Cole et al., 1998 (Nature 393 537-544) et contenant seq39A

*[The following sequence was obtained from a DNA microsequencing experiment.]*

1/1  
ATG TGG ACA ACG CGG TTG GTT CGA TCC GGA CTC GCC GCG CTG TGC GCA GTA CTG GTA  
Met trp thr thr arg leu val arg ser gly leu ala ala leu cys ala ala val leu val  
61/21  
TCG AGC GGC TGC GCA CGG TTC AAC GAC GCT CAA TCT CAG CCG TTC ACC ACC GAA CCG GAG  
ser ser gly cys ala arg phe asn asp ala gln ser gln pro phe thr thr glu pro glu  
121/41  
CTG CGG CCC CAA CCC AGC TCG ACA CCT CCC CCC CGG CCG CTG CCG CCG GTT CCC TTT  
leu arg pro gln pro ser ser thr pro pro pro pro pro pro pro leu pro pro val pro phe  
181/61  
CCC AAG GAA TGT CCG GCG CCG GGC GTG ATG CAA GGC TGC CTT GAG AGC ACC AGC GGC TTG  
pro lys glu cys pro ala pro gly val met gln gly cys leu glu ser thr ser gly leu  
241/91  
ATC ATG GGC ATC GAC AGC AAG ACC GCA CTG GTC GCC GAG GCG ATC ACC GGT GCC GTC GAG  
ile met gly ile asp ser lys thr ala leu val ala gln arg ile thr gly ala val glu  
301/101  
GAG ATC TGT ATC AGC GCG GAG CCG AAG GTA AAG ACB GTC ATC CCC GTG CAT CCT GCC GGT  
glu ile ser ile ser ala glu pro lys val lys thr val ile pro val asp pro ala gly  
361/121  
GAC GGT GGC TTG ATG GAC ATT GTG CTG TCG CCC ACC TAG TGG CAA GAC CCG CTG ATG TAC  
asp gly gly leu met asp ile val leu ser pro thr tyr ser gln asp arg leu met tyr  
421/141  
GCC TAC ATC AGC ACS CCC ACC GAC AAC CCG GTG GTG CGA GTS CCC GAC GGC GAC ATC CCC  
ala tyr ile ser thr pro thr asp asn arg val val arg val ala asp gly asp ile pro  
481/161  
AAG GAC ATC CTG ACC GGC ATC CCC AAA GST GCT GGC GGT AAC ACC GGG GCG CTG ATC TTC  
lys asp ile leu thr gly ile pro lys gly ala ala gly asn thr gly ala leu ile phe  
541/191  
ACC AGT CCC ACC ACS CTG GTC GTG ATG ACC GGG GAT TTT GGS GAG GCS GCG TTG GGC GGC  
thr ser pro thr thr leu val val met thr gly asp ala gly asp pro ala leu ala ala  
601/211  
GAT CBC CAA TCG TTG GCG GST AAG GTC CTG CST ATC GAA CA CCH AS ACC ATC GGC CAG  
asp pro gln ser leu ala gly lys val leu arg ile ala gln pro thr thr ile gly gin  
661/221  
ACG CSG CSB ACS ADS GCG CTG TOT GGC ATC GTC TST SPT ST ST ST ST ST ATC GAT CCS  
thr pro pro thr thr ala leu ser gly ile gly ser sly sly sly sly sly sly sly sly sly  
721/241  
GTC GAC GAD TOS GTA TAT GTC GUC SAC CSC ACG GTA A A A A GA TA ITT CAS CWS ATP  
val asp gly ser leu tyr val ala asp arg thr pro thr ala asp arg leu gin arg ile  
781/261  
AAC AAG AAC TDS GAS GTC TOT ACG GTA THS AHS THS SPS ASD AAH SHS DSH DHS GHG  
thr lys asn ser glo val ser thr val thr thr thr thr thr thr thr thr thr thr thr thr thr  
841/291  
TOT GCD GSX AIS SAS GGL ASM STD STN STS AAS ITS IIS IAS IBS IDS ISS ISL ISM ISV ISW  
lys ala ile pro ipr gly thr val leu val leu leu leu leu leu leu leu leu leu leu leu leu  
901/311  
STS SGS STP STQ STA STE ASL ASE AST ATS AUS AVS AWS AWY AXS AYU AYV AZS AZW BLS  
thr  
961/331  
AST SAT SLI SAT SIG STR SMA STM STN STP STS STT STV STW STY SVR SWR SWY SXD SYD SZD  
thr his thr thr thr tip ala thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr  
1021/351  
*(Sequence continues beyond visible range)*

[illegible]

127/185

961/321  
 GTC TCT ACG GTA TGG ACC TGG CCG GAC AAG CCC GGC GTG GCC GGG TGT GCC GCG ATG GAC  
 val ser thr val trp thr trp pro asp lys pro gly val ala gly cys ala ala met asp  
 1021/341  
 GGC ACC GTG CTG GTC AAC CTG ATT AAT ACC AAA CTG ACG GTG GCG GTC CCG CTC GCG CCG  
 gly thr val leu val asn leu ile asn thr lys leu thr val ala val arg leu ala pro  
 1081/361  
 TCG ACC GGT GCG GTC ACC GGA GAA CCC GAC GTT GTC CCG AAA GAC ACT CAT GCG CAT GCG  
 ser thr gly ala val thr gly glu pro asp val val arg lys asp thr his ala his ala  
 1141/381  
 TGG GCA TTA CCG ATG TCG CCG GAC GGC AAC GTG TGG GGA GCC ACC GTC AAC AAG ACC GCC  
 trp ala leu arg met ser pro asp gly asn val trp gly ala thr val asn lys thr ala  
 1201/401  
 GGC GAC GCC GAG AAG CTC GAC GAT GTG GTG TTC CCG CTC TTC CCG CAG GGT GGC GGC TTC  
 gly asp ala glu lys leu asp asp val val phe pro leu phe pro gin gly gly gly phe  
 1261/421  
 CCG CGT AAC AAC GAC GAC AAG ACC TGA  
 pro arg asn asn asp asp lys thr opa

SEQ ID N° 39F (suite)

FIGURE 39F (suite)

1/1  
 GAA GGC GTT GTT GAG CCG GCG CAC GAA AAC GAT CGT TGT GTG TAC ATT GGT GTG TAT GGC  
 glu gly leu val glu pro ala his glu asn asp arg cys val tyr ile gly val tyr gly  
 61/21  
 TCG GTT GAA CGT GTA TGT GCG CGA CGA ATT GGT GGA GTT CCG CAG GGC GCG GCG GTT GAA  
 ser val glu arg val cys ala arg arg ile gly gly ala arg glu gly ala gly leu ala  
 121/41  
 CGT CTC GGC GGT GAC TCA GCG CGG GAT CAG TGG CGA GTT GCA GAA CTC CCG AAC CGA TTT  
 arg leu gly ala asp ser gly arg asp ala tyr arg val gly ala leu arg asn arg cys  
 181/61  
 GCG GTT TGA GCG GTT GCA AAT GAG AAG CAG CCG CGG TGG GCA TGA TGA GGT GAT GAT TGT  
 val ala opa gly val gly thr ala lys ala arg his ser ala ala opa arg ala gly cys  
 241/81  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 301/101  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 361/121  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 421/141  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 481/161  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 541/181  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 601/201  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 661/221  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 721/241  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 781/261  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 841/281  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 901/301  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 961/321  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 1021/341  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 1081/361  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 1141/381  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 1201/401  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat  
 1261/421  
 GAT GCA TAT GTT TGT GCA TGA GTT GAA AAT GTG AAG GCA TTT GAT TGT GAT GAT GAT GAT  
 asp ala thr thr thr ala tga gtt gaa aat gtg aag gca ttt gat tgt gat gat gat gat gat

128/185

```

1/1          31/11
AAG GCC TTG TTG AGC CGG CGC ACG AAA ACG ATC GTT GTG TGT ACA TTG GTG TGT ATG GCT
lys ala leu leu ser arg arg thr lys thr ile val val cys thr leu val cys met ala
61/21       91/31
CGG TTG AAC GTG TAT GTG CCC GAC GAA TTG GCG GAG CGC GCC AGG GCG CGG GGC TTG AAC
arg leu asn val tyr val pro asp glu leu ala glg arg ala arg ala arg gly leu asn
121/41      151/51
GTC TCG GCG CTG ACT CAG GCC GCG ATC AGT GCC GAG TTG GAG AAC TCC GCA ACC GAT GCG
val ser ala leu thr gln ala ala ile ser ala glu leu glu asn ser ala thr asp ala
181/61      211/71
TGG CTT GAG GGG TTG GAA CCC AGA AGC ACC GGC GCT CGG CAT GAT GAC GTG CTG GGT GCG
trp leu glu gly leu glu pro arg ser thr gly ala arg his asp asp val leu gly ala
241/81      271/91
ATC GAT GCC GCT CGC GAT GAG TTC GAA GCG TGA GAG CAT CGC CCA CTT CGC CGG CGG ACC
ile asp ala ala arg asp glu phe glu ala GBA glu his arg pro leu arg arg arg ser
301/101     331/111
AGG TGG TCG TCG ACG CGA GTG CCA TGG TGG ATC
arg trp ser ser thr arg val pro trp trp ile

```

SEQ ID N° 40B

FIGURE 40B

```

1/1
AGG CCT TGT TGA GCG GCG GCA CGA AAA CGA TCG TTG TGT GTA CAT TGG TGT GTA TGG CTC
arg pro cys oPa ala gly ala arg lys arg ser leu cys val his trp cys val trp leu
61/21
GCT TGA AGG TGT ATG TGC CCG ACG AAT TCG CCG AGG CCG CGA GCG CCG GCG GCT TGA AGG
gly oPa thr cys met cys pro thr asn trp arg ser ala pro gly arg gly ala oPa thr
121/41
TCT CGG CCG TGA CTC AGG CCG CGA TGA GTS CCG AGT TGG AGA ACT CCG CAA CGG AAG GGT
ser arg arg oPa leu arg pro ala ser val his ser his arg thr pro ala pro met arg
141/61
GGT TTS AGG GGT TGG AAA GGA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA GAA
gly leu arg gly trp ala pro ala ala ala ala ala ala ala ala ala ala ala ala ala ala
161/81
TGG ATG TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT
leu thr pro leu ala met ser ser cys arg gly ser thr ala his pro ala ala gly ala
181/101
GCT GGT GGT GAA TTT TAT TGC TAT TTT TTT TTT
gly gly ala met thr thr thr his thr thr

```



129/185

Séquence codante Rv0549c prédite par Cole et al., 1998 (Nature 393:537-544)et contenant seq40A

```

1/1                               31/11
gtg aga gca tgc ccc act tgc ccg ccg gag cag gtg gtc gtc gac gcg agt gcc atg gtg
val arg ala ser pro thr ser pro pro glu gln val val val asp ala ser ala met val
61/21                               91/31
gat cta ctg gct cgc act agc gat ccg tgc tct gcg gtg cgc gcg ccg ctg gct cgg acc
asp leu leu ala arg thr ser asp arg cys ser ala val arg ala arg leu ala arg thr
121/41                               151/51
gcg atg cac gcg ccg gcg cac ttc gat gca gag gtg ttg tgc gcg ctg gcg cgc atg cag
ala met his ala pro ala his phe asp ala glu val leu ser ala leu gly arg met gln
181/61                               211/71
cgc gcc gcc gca ctc acc gtt gcc tat gtc gat gcg gca ctg gag gag ttg cga cag gtg
arg ala gly ala leu thr val ala tyr val asp ala ala leu glu glu leu arg gln val
241/81                               271/91
ccg gtg act cga cac ggt ctt tgc tgc ctg ctt ggt gga gcg tgg tgc cgc cgc gac acc
pro val thr arg his gly leu ser ser leu leu ala gly ala trp ser arg arg asp thr
301/101                               331/111
ctc cgc ctg acc gat gcc ctc tac gtc gag ctg ggc gaa agc gca ggt ctg gtg ttg ttg
leu arg leu thr asp ala leu tyr val glu leu ala glu thr ala gly leu val leu leu
361/121                               391/131
acc acc gac gaa aga ttg gca cgc gcc tgg cca tgg gtt cca gcc atc gcc tga
thr thr asp glu arg leu ala arg ala trp pro ser ala his ala ile gly opa

```

SEQ ID N° 40D

FIGURE 40D

CRF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv0549c

```

1/1                               31/11
tga gtc cga agt gtg aga gca tgc ccc act tgc ccg ccg gag cag gtg gtc gtc gac gcg
CPA val arg ser val arg ala ser pro thr ser pro pro glu gln val val val asp ala
61/21                               91/31
agt gcc atg gtg gat cta ctg gct cgc act agc gat ccg tgc tct gcg gtg cgc gcg ccg ctg gct cgg acc
ser ala met val asp leu leu ala arg thr ser asp arg cys ser ala val arg ala arg thr
121/41                               151/51
gcg atg cac gcg ccg gcg cac ttc gat gca gag gtg ttg tgc gcg ctg gcg cgc atg cag
ala met his ala pro ala his phe asp ala glu val leu ser ala leu gly arg met gln
181/61                               211/71
cgc gcc gcc gca ctc acc gtt gcc tat gtc gat gcg gca ctg gag gag ttg cga cag gtg
arg ala gly ala leu thr val ala tyr val asp ala ala leu glu glu leu arg gln val
241/81                               271/91
ccg gtg act cga cac ggt ctt tgc tgc ctg ctt ggt gga gcg tgg tgc cgc cgc gac acc
pro val thr arg his gly leu ser ser leu leu ala gly ala trp ser arg arg asp thr
301/101                               331/111
ctc cgc ctg acc gat gcc ctc tac gtc gag ctg ggc gaa agc gca ggt ctg gtg ttg ttg
leu arg leu thr asp ala leu tyr val glu leu ala glu thr ala gly leu val leu leu
361/121                               391/131
acc acc gac gaa aga ttg gca cgc gcc tgg cca tgg gtt cca gcc atc gcc tga
thr thr asp glu arg leu ala arg ala trp pro ser ala his ala ile gly opa

```

SEQ ID N° 40E

FIGURE 40E

130/185

```

1/1                               31/11
CCT GGC CGG GAC GCC TAC GTG TAG CCC GCG GCT AGC ACA GGA TAG CCA TTG TTG TGC GGT
pro gly arg asp ala tyr val AMB pro ala ala ser thr gly AMB pro leu leu cys gly
61/21                               91/31
AGC GCC AAA ACG ATC AGC CCT TCG CGG ACA TGT CAG CAC CCG CCT TGG CCG GGA GAG CGG
ser ala lys thr ile ser pro ser arg thr cys gln his pro pro trp pro gly glu arg
121/41                               151/51
CGT CGT GAC CGT GCT GTC ACC ACG TCT GGT TAG GCT CGG GGC GCG GGC TGG CGC GGA GGA
arg arg asp arg ala val thr thr ser gly AMB ala arg gly ala gly trp arg gly gly
181/61                               211/71
GGT GTG TTG CGG AGG AGG TGT GTT GTA GTG GGG ACG GCG GAT CGG CCG TTG GAC GCC TCG
gly val leu arg arg arg cys val val val gly thr ala asp arg pro leu asp ala ser
241/81                               271/91
GCC TTG CGG GAC TGG GCA CAC GCC GTC GTC AGC GAT C
ala leu arg asp trp ala his ala val val ser asp

```

SEQ ID N° 41A

FIGURE 41A

```

1/1                               31/11
CTG GGC GGG AGG CCT AGG TGT AGC CCG CGG GGA GGA TAG GAT AAT CAT TGT TGT GCG GTA
leu ala gly thr pro thr cys ser pro arg leu ala ala asp ser his cys cys ala val
61/21                               91/31
GCG CCA AAA CGA TGA GGC CTT CGG GGA CAT TGT AGC AAT CAT CTT GGT CGG GAG AAT GGC
ala pro lys arg ser ala leu arg gly his val ser thr arg leu gly arg glu ser gly
121/41                               151/51
GTC GTG ACC GTG CTG TGA CCA CCG CTG GTT AGG GTC GCG GTC GCG GGT GGT GCG GAG GAG
val val thr val leu ser pro arg leu val arg leu gly ala arg ala gly ala glu glu
181/61                               211/71
GTG TGT TGT GGA GGA GAT GTA TTT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT TAT
val cys cys gly gly gly val leu AMB asp gly thr thr thr thr thr thr thr thr thr
241/81                               271/91
GAT TGT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT GAT
thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

```

SEQ ID N° 41B

FIGURE 41B

131/185

```

1/1                               31/11
TGG CCG GGA CGC CTA CGT GTA GCC CGC GGC TAG CAC AGG ATA GCC ATT GTT GTG CGG TAG
trp pro gly arg leu arg val ala arg gly AMB his arg ile ala ile val val arg AMB
61/21                               91/31
CGC CAA AAC GAT CAG CCC TTC GCG GAC ATG TCA GCA CCC GCC TTG GCC GGG AGA GCG GCG
arg gln asn asp gln pro phe ala asp met ser ala pro ala leu ala gly arg ala ala
121/41                               151/51
TCG TGA CCG TGC TGT CAC CAC GTC TGG TTA GGC TCG GGG CGC GGG CTG SCG CCG AGG AGG
ser OPA pro cys cys his his val trp leu gly ser gly arg gly leu ala arg arg arg
181/61                               211/71
TGT GTT CCG GAG GAG GTG TGT TGT AGT GGG GAC GGC GGA TCG GCC GTT GGA CGC CTC GGC
cys val ala glu glu val cys cys ser gly asp gly ser ala val gly arg leu gly
241/81                               271/91
CTT CCG GGA CTG GGC ACA CGC CGT CGT CAG CGA TC
leu ala gly leu gly thr arg arg arg gln arg

```

SEQ ID N° 41C

FIGURE 41C

Sequence codante kv29750 prédite par Gale et al, 1998 (Nature 393: 537-544) et contenant seq41A

```

1/1                               31/11
GAG GAG GAG GAG GAT GGG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG GAG
val gly thr ala asp arg pro leu asp ala ser ala leu arg asp trp ala his ala val
61/21                               91/31
GTC GTC GAT CTG ATC CTA GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
val ser asp leu ile leu his ile ser ala ile asp asp leu asp val phe pro val ala
121/41                               151/51
GTC GTC GAT CTG ATC CTA GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
val ser asp leu ile leu his ile ser ala ile asp asp leu asp val phe pro val ala
181/61                               211/71
GTC GTC GAT CTG ATC CTA GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
val ser asp leu ile leu his ile ser ala ile asp asp leu asp val phe pro val ala
241/81                               271/91
GTC GTC GAT CTG ATC CTA GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC GTC
val ser asp leu ile leu his ile ser ala ile asp asp leu asp val phe pro val ala

```

132/185

ORF d'après Cole et al, 1998 (Nature 393: 537-544) et contenant Rv2975c

```

1/1                               31/11
tag gct cgg ggc gcg ggc tgg cgc gga gga ggt gtg ttg cgg agg agg tgt gtt gta gtg
AMB ala arg gly ala gly trp arg gly gly gly val leu arg arg arg cys val val val
61/21                               91/31
ggg acg gcg gat cgg ccg ttg gac gcc tgc gcc ttg cgg gac tgg gca cac gcc gtc gtc
gly thr ala asp arg pro leu asp ala ser ala leu arg asp trp ala his ala val val
121/41                               151/51
agc gat ctg atc ctc cac atc gac gag atc aac cgg ctc aat gtg ttc ccg gtc gct gac
ser asp leu ile leu his ile asp glu ile asn arg leu asn val phe pro val ala asp
181/61                               211/71
tcc gat acc ggc gtc aac atg ctg ttc acc atg cgt gcc gcg gtc gta gaa gct gat ttg
ser asp thr gly val asn met leu phe thr met arg ala ala val val glu ala asp leu
241/81                               271/91
cac gcg aat tgc cag gct gac gcc gaa gac gtg gcg cgg gtt gcg gcc gct ctc gcg gcc
his ala asn ser gln ala asp ala glu asp val ala arg val ala ala ala leu ala ala
301/101
ggc gcg cgt tga
gly ala arg opa

```

SEQ ID N° 41F

FIGURE 41F

séquence Rv 2974C prédite par Cole et al. (Nature 393:537-544) et pouvant être dans la même phase de lecture que Seq41B. Le séquençage de cette région fait apparaître dans un cas sur trois une déletion de deux nucléotides mettant en phase bserve dans

```

1/1                               31/11
ttg aac gga gct cgc ggc aac tcc ggc gty atc cgt tcc tag atc ctg cgg ggg atc gaa
leu asn gly ala arg gly asn ser gly val ile leu ser ala ile leu arg gly ile ala
61/21                               91/31
aaa gta aat ggc att ggc gaa ggc att tcc aac ggc att tcc aac ggc att tcc aac ggc
ala val thr ala thr ala ala ala ala ala ser gly ala val leu arg ala val asp ala asp
121/41                               151/51
tcc gta aat ggc att ggc gaa ggc att tcc aac ggc att tcc aac ggc att tcc aac ggc
ala leu gly ala ala leu thr arg ala ala ala leu thr ala ala leu thr ala ala leu
181/61                               211/71
ttg gta aat ggc att ggc gaa ggc att tcc aac ggc att tcc aac ggc att tcc aac ggc
ala val pro gly thr ala val asp val leu arg ala ala ala ggc ala val asp ala ggc
241/81                               271/91
aaa gta aat ggc att ggc gaa ggc att tcc aac ggc att tcc aac ggc att tcc aac ggc
ala val thr ala thr ala ala ala ala ala ser gly ala val leu arg ala val asp ala
301/101
ggc gcg cgt tga
gly ala arg opa

```

SEQ ID N° 41F

FIGURE 41F

La séquence de nucléotides présentée ci-dessus est une séquence de nucléotides

[illegible]

134/185

Seq41T comprenant seq 41F et seq 41S

1/1

31/11

tta ggc tgc ggg cgc ggg ctg ggc cgg agg agg tgt gtt ggc gag gag gtg tgt tgt agt  
 leu gly ser gly arg gly leu ala arg arg arg cys val ala glu glu val cys cys ser  
 AMB ala arg gly ala gly trp arg gly gly val leu arg arg arg cys val val val  
 arg leu gly ala arg ala gly ala glu glu val cys cys gly gly gly val leu AMB trp

61/21

91/31

ggg gac ggc gga tgc gcc gtt gga cgc ctc gcc ctt ggc gga ctg ggc aca cgc cgt cgt  
 gly asp gly gly ser ala val gly arg leu gly leu ala gly leu gly thr arg arg arg  
 gly thr ala asp arg pro leu asp ala ser ala leu arg asp trp ala his ala val val  
 gly arg arg ile gly arg trp thr pro arg pro cys gly thr gly his thr pro ser ser

121/41

151/51

cag cga tct gat cct cca cat cga cga gat caa ccg gct caa tgt gtt ccc ggt cgc tga  
 gln arg ser asp pro pro his arg arg asp gln pro ala gln cys val pro gly arg OPA  
 ser asp leu ile leu his ile asp glu ile asn arg leu asn val phe pro val ala asp  
 ala ile OPA ser ser thr ser thr arg ser thr gly ser met cys ser arg ser leu thr

181/61

211/71

ctc cga tac cgg cgt caa cat gct gtt cac cat ggc tgc cgc ggt cgt aga agc tga ttt  
 leu arg tyr arg arg gln his ala val his his ala cys arg gly arg arg ser OPA phe  
 ser asp thr gly val asn met leu phe thr met arg ala ala val val glu ala asp leu  
 pro ile pro ala ser thr cys cys ser pro cys val pro arg ser AMB lys leu ile cys

241/81

271/91

gca cgc gaa ttc gca ggc tga cgc cga aga cgt ggc ggc ggt tgc ggc cgc tct cgc ggc  
 ala arg glu phe ala gly OPA arg arg arg arg gly ala gly cys gly arg ser arg gly  
 his ala asn ser gln ala asp ala glu asp val ala arg val ala ala ala leu ala ala  
 thr arg ile arg arg leu thr pro lys thr trp arg gly leu arg pro leu ser arg pro

301/101

331/111

cgg cgc ggc ttg aac gga gct cgc ggc aac tcc ggc gtc atc ctg tcc cag atc ctg cgc  
 arg arg ala leu asn gly ala arg gly asn ser gly val ile leu ser gln ile leu arg  
 gly ala arg OPA thr glu leu ala ala thr pro ala OPA ser cys pro arg ser cys ala  
 ala arg val glu arg ser ser arg gln leu arg arg asp pro val pro asp pro ala arg

361/121

391/131

ggg atc gca gag gtg acc ggc act ggc gcc gcc gcc trt ggc ggc gta ttg cgg ggc gtc  
 gly ile ala glu val thr ala thr ala ala ala ala ser gly ala val leu arg ala val  
 gly ser gln arg OPA pro arg leu arg pro pro pro leu ala arg tyr cys gly arg ser  
 asp arg arg gly asp arg asp cys gly arg arg leu trp arg gly ile ala gly gly arg

421/141

451/151

gac gcc aac gcc ctc ggc gcc gcc ttc tag gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 asp ala asn ala leu gly ala ala leu tip arg gly val glu leu val val ala ser met  
 thr pro thr pro ser gly pro arg cys gly ala ala ser ser tip ser ser arg arg trp  
 arg gln arg pro arg gly arg val val ala arg arg arg val gly arg arg val asp gly

481/161

511/171

ggt gcc gtc cag gtc cgc gcc act atc gcc tag gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 gly gly val glu val pro gly thr ile val ser val leu ser ala ala ala gly ala val  
 val ala trp arg cys arg glu leu ser ser arg cys cys gly pro pro pro glu pro ser  
 trp arg gly arg ala gly asn tyr arg leu gly ala cys gly arg ala ser arg arg

541/181

571/191

ggt gcc gtc cag gtc cgc gcc act atc gcc tag gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 trp ala cys ala met gly leu ala gly ala val thr ala ala gly arg ala ala val  
 thr ser ala arg thr arg gly thr pro val arg ser pro pro pro val thr arg arg ser  
 pro val arg ala arg gly val gly arg cys gly arg arg arg ala ala arg gly gly pro

601/201

631/211

ggt gcc gtc cag gtc cgc gcc act atc gcc tag gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 ile ala leu ala gly thr arg arg arg arg arg arg arg arg arg arg arg arg arg  
 arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg  
 arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg

661/221

691/231

721/241  
gca cct gcc cgg gcg gtc tac gaa ccc tgg ccg cgc gcg ttg ccg acc gac acg gct acc  
ala pro ala arg ala val tyr glu pro ser pro arg ala leu pro thr asp thr ala thr  
his leu pro gly arg ser thr asn pro arg arg ala arg cys arg pro thr arg leu pro  
thr cys pro gly gly leu arg thr leu ala ala arg val ala asp arg his gly tyr pro  
781/261 811/271  
caa cgc ccc gcc ccg caa ttc gag gtg atg tat ctg ttg gcg gta tgt gat gct gca gcg  
gln arg pro ala pro gln phe glu val met tyr leu leu ala val cys asp ala ala ala  
asn ala pro pro arg asn ser arg opa cys ile cys trp arg tyr val met leu gln arg  
thr pro arg pro ala ile arg gly asp val ser val gly gly met opa cys cys ser gly  
841/281 871/291  
gcg gac cag ttg ccg gat cga ctc aag gaa ttg ggt gag tgg gtg gcc atc gcc gct gct  
ala asp gln leu arg asp arg leu lys glu leu gly glu ser val ala ile ala ala ala  
arg thr ser cys gly ile asp ser arg asn trp val ser arg trp pro ser pro leu leu  
gly pro val ala gly ser thr gln gly ile gly opa val gly gly his arg arg cys ser  
901/301 931/311  
ccg ccc gac agc tac tcc gla ccc gtc cag acc gac gac gcc ggt gcc gcc gtg gaa gcc  
pro pro asp ser tyr ser val his val his thr asp asp ala gly ala ala val glu ala  
arg pro thr ala thr pro tyr thr ser thr pro thr thr pro val pro pro trp lys pro  
ala arg gln leu leu arg thr arg pro his arg arg atg cys arg arg gly ser arg  
961/321 991/331  
gga ttg gcg gtg ggg cga gtt agc cgg atc gtg atc tgg gcg ctc ggt tcc ggg acc agc  
gly leu ala val gly arg val ser arg ile val ile ser ala leu gly ser gly thr ser  
asp trp arg trp gly glu leu ala gly ser opa ser arg arg ser val pro gly pro ala  
ile gly gly gly ala ser amb pro asp arg asp leu gly ala arg phe arg asp gln arg  
1021/341 1051/351  
gga ttg ccg gcc ggt gcc tgg acg cgg gcc cgc gcc gtg ctg gcg gtc gtc gac gcc gac  
gly leu pro ala gly gly trp thr arg gly arg ala ala val leu ala val val asp gly asp  
asp cys arg pro val ala gly arg gly ala ala pro cys trp arg ser ser thr ala thr  
ile ala gly arg trp leu asp ala gly pro arg arg ala gly gly arg arg arg arg arg  
1061/361 1111/371  
ggt gcc gcc gag ctg ttc gcc ggy gag gcc gcc tgc gtg ctg cga ccg ggt cca gac gcc  
gly ala ala glu leu phe ala gly glu gly ala cys val leu arg pro gly pro asp ala  
val pro pro ser cys ser pro gly arg ala pro ala cys cys asp arg val gln thr pro  
cys arg arg ala val arg arg gly gly arg leu arg ala ala thr gly ser arg arg arg  
1141/381 1171/391  
gtg aca ccg gcc gcc gat atc agt gcc cac cag ctg gtg cgg gcc gtg gta gac acc gcc  
val thr pro ala ala asp ile ser ala his gln leu val arg ala val val asp thr gly  
opa his arg pro pro ile ser val pro thr ser trp cys gly pro trp amb thr pro ala  
asp thr gly arg arg tyr gln cys pro pro ala gly ala gly arg gly arg his arg arg  
1201/401 1231/411  
ggt gcg cac gtg atg gtg ctg ccc aat gcc tat gtg gcc gcc caa caa cta gtc gtc gcc  
ala ala his val met val leu pro asn gly tyr val ala ala gln glu leu val ala gly  
pro arg thr opa trp cys cys pro met ala met trp trp pro lys asn trp trp pro gly  
arg ala asp gly ala ala gln trp leu cys gly arg arg arg thr gly gly arg val  
1241/421 1291/441  
tat acc ggc arg gtc gca tgc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc  
ggt thr ala ala gly trp gly val asp val val pro val pro thr gly ser met val  
val pro arg met met arg gly ala ser thr trp trp cys arg trp arg arg trp arg  
tyr arg gly asp arg leu gly arg arg arg gly thr arg ala asp thr ala asp gly ala  
1301/441 1351/451  
gag ggt ttc ggc ggt gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc  
ala gly leu ala ala leu ala val his asp ala ala ala ala ala val asp arg gly arg  
arg gly trp pro arg trp pro val met thr arg ala ala ala ala ala ala ala ala ala  
gly val asp ala gly arg ala ala ala arg gly pro arg arg arg arg arg arg arg arg  
1361/451 1411/461

136/185

1441/481  
gcg ctg acc tgg gcc ggt acc tgc aag cgg ggc gac ggt ctg ggt atc gcg ggc gac gag  
ala leu thr trp ala gly thr cys lys pro gly asp gly leu gly ile ala gly asp glu  
arg OPA pro gly pro val pro ala ser arg ala thr val trp val ser arg ala thr arg  
ala asp leu gly arg tyr leu gln ala gly arg arg ser gly tyr arg gly arg arg gly  
1501/501  
gtg ctg atc gtc gcc gac gat gtc gcc gcg gcg gcc atc ggt ctg gtc gac ctg ttg ttg  
val leu ile val ala asp asp val ala ala ala ile gly leu val asp leu leu leu  
cys OPA ser ser pro thr met ser pro arg arg pro ser val trp ser thr cys cys trp  
ala asp arg arg arg arg cys arg arg gly gly his arg ser gly arg pro val val gly  
1561/521  
gca tgg gga gcc gat ctg gtg acg ctg cta att gcc gcc ggc gta acc gaa gac gtg gct  
ala ser gly gly asp leu val thr val leu ile gly ala gly val thr glu asp val ala  
his arg glu ala ile trp OPA arg cys OCH leu ala pro ala OCH pro lys thr trp leu  
ile gly arg arg ser gly asp gly ala asn trp arg arg arg asn arg arg arg gly cys  
1621/541  
gtc gtc ctg gaa cgg cat ctg cac gac cac cat cca gcc acc gag ctg gtc tcc tac cgc  
val val leu glu arg his val his asp his his pro gly thr glu leu val ser tyr arg  
ser ser trp asn gly met cys thr thr thr ile gln ala pro ser trp ser pro thr ala  
arg pro gly thr ala cys ala arg pro pro ser arg his arg ala gly leu leu pro his  
1681/561  
acc gga cac cgc gcc gac gcg ctg ctg atc ggg gtc gag tag  
thr gly his arg gly asp ala leu leu ile gly val glu AMR  
pro asp thr ala ala thr arg cys OPA ser gly ser ser  
arg thr pro arg arg arg ala ala asp arg gly arg val;

SEQ ID N° 41T (suite 2)

FIGURE 41T (suite 2)

[illegible]

1000

1997, 1998, 1999, 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023, 2024, 2025, 2026, 2027, 2028, 2029, 2030, 2031, 2032, 2033, 2034, 2035, 2036, 2037, 2038, 2039, 2040, 2041, 2042, 2043, 2044, 2045, 2046, 2047, 2048, 2049, 2050, 2051, 2052, 2053, 2054, 2055, 2056, 2057, 2058, 2059, 2060, 2061, 2062, 2063, 2064, 2065, 2066, 2067, 2068, 2069, 2070, 2071, 2072, 2073, 2074, 2075, 2076, 2077, 2078, 2079, 2080, 2081, 2082, 2083, 2084, 2085, 2086, 2087, 2088, 2089, 2090, 2091, 2092, 2093, 2094, 2095, 2096, 2097, 2098, 2099, 2100, 2101, 2102, 2103, 2104, 2105, 2106, 2107, 2108, 2109, 2110, 2111, 2112, 2113, 2114, 2115, 2116, 2117, 2118, 2119, 2120, 2121, 2122, 2123, 2124, 2125, 2126, 2127, 2128, 2129, 2130, 2131, 2132, 2133, 2134, 2135, 2136, 2137, 2138, 2139, 2140, 2141, 2142, 2143, 2144, 2145, 2146, 2147, 2148, 2149, 2150, 2151, 2152, 2153, 2154, 2155, 2156, 2157, 2158, 2159, 2160, 2161, 2162, 2163, 2164, 2165, 2166, 2167, 2168, 2169, 2170, 2171, 2172, 2173, 2174, 2175, 2176, 2177, 2178, 2179, 2180, 2181, 2182, 2183, 2184, 2185, 2186, 2187, 2188, 2189, 2190, 2191, 2192, 2193, 2194, 2195, 2196, 2197, 2198, 2199, 2200, 2201, 2202, 2203, 2204, 2205, 2206, 2207, 2208, 2209, 2210, 2211, 2212, 2213, 2214, 2215, 2216, 2217, 2218, 2219, 2220, 2221, 2222, 2223, 2224, 2225, 2226, 2227, 2228, 2229, 2230, 2231, 2232, 2233, 2234, 2235, 2236, 2237, 2238, 2239, 2240, 2241, 2242, 2243, 2244, 2245, 2246, 2247, 2248, 2249, 2250, 2251, 2252, 2253, 2254, 2255, 2256, 2257, 2258, 2259, 2260, 2261, 2262, 2263, 2264, 2265, 2266, 2267, 2268, 2269, 2270, 2271, 2272, 2273, 2274, 2275, 2276, 2277, 2278, 2279, 2280, 2281, 2282, 2283, 2284, 2285, 2286, 2287, 2288, 2289, 2290, 2291, 2292, 2293, 2294, 2295, 2296, 2297, 2298, 2299, 2300, 2301, 2302, 2303, 2304, 2305, 2306, 2307, 2308, 2309, 2310, 2311, 2312, 2313, 2314, 2315, 2316, 2317, 2318, 2319, 2320, 2321, 2322, 2323, 2324, 2325, 2326, 2327, 2328, 2329, 2330, 2331, 2332, 2333, 2334, 2335, 2336, 2337, 2338, 2339, 2340, 2341, 2342, 2343, 2344, 2345, 2346, 2347, 2348, 2349, 2350, 2351, 2352, 2353, 2354, 2355, 2356, 2357, 2358, 2359, 2360, 2361, 2362, 2363, 2364, 2365, 2366, 2367, 2368, 2369, 2370, 2371, 2372, 2373, 2374, 2375, 2376, 2377, 2378, 2379, 2380, 2381, 2382, 2383, 2384, 2385, 2386, 2387, 2388, 2389, 2390, 2391, 2392, 2393, 2394, 2395, 2396, 2397, 2398, 2399, 2400, 2401, 2402, 2403, 2404, 2405, 2406, 2407, 2408, 2409, 2410, 2411, 2412, 2413, 2414, 2415, 2416, 2417, 2418, 2419, 2420, 2421, 2422, 2423, 2424, 2425, 2426, 2427, 2428, 2429, 2430, 2431, 2432, 2433, 2434, 2435, 2436, 2437, 2438, 2439, 2440, 2441, 2442, 2443, 2444, 2445, 2446, 2447, 2448, 2449, 2450, 2451, 2452, 2453, 2454, 2455, 2456, 2457, 2458, 2459, 2460, 2461, 2462, 2463, 2464, 2465, 2466, 2467, 2468, 2469, 2470, 2471, 2472, 2473, 2474, 2475, 2476, 2477, 2478, 2479, 2480, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2608, 2609, 2610, 2611, 2612, 2613, 2614, 2615, 2616, 2617, 2618, 2619, 2620, 2621, 2622, 2623, 2624, 2625, 2626, 2627, 2628, 2629, 2630, 2631, 2632, 2633, 2634, 2635, 2636, 2637, 2638, 2639, 2640, 2641, 2642, 2643, 2644, 2645, 2646, 2647, 2648, 2649, 2650, 2651, 2652, 2653, 2654, 2655, 2656, 2657, 2658, 2659, 2660, 2661, 2662, 2663, 2664, 2665, 2666, 2667, 2668, 2669, 2670, 2671, 2672, 2673, 2674, 2675, 2676, 2677, 2678, 26

1. *Journal of the American Medical Association*, 1997; 277: 1027-1031.



137/185

[illegible]

SEQ ID N° 42B

FIGURE 42B

[illegible]

**ACKNOWLEDGMENTS**

THE UNIVERSITY OF CHICAGO PRESS

138/185

Séquence codante Rv2622 prédite par Cole et al., 1998 (Nature 393:537-544) et contenant seq42A:

```

1/1                               31/11
atg gcc aac aaa cgt ggc aat gcc ggg cag cct ctg ccc ttg tcg gat cga gac gac gac
Met ala asn lys arg gly asn ala gly gln pro leu pro leu ser asp arg asp asp asp
61/21                               91/31
cac atg cag ggg cac tgg ctg ctg gcc cgg ctg gcc aag cgg gtg ctg cgt ccc gcc gcc
his met gln gly his trp leu leu ala arg leu gly lys arg val leu arg pro gly gly
121/41                               151/51
gtc gaa ctc acc cgg aca ctg ctg gcc cgc gcc gag gtg acc gac gcc gac gtg ctc gag
val glu leu thr arg thr leu leu ala arg ala glu val thr asp ala asp val leu glu
181/61                               211/71
ctg gca cgg ggc ctg ggc cgc acc gca gcc gaa atc ttg gcc cgc aac ccg cgg tcg tac
leu ala pro gly leu gly arg thr ala ala glu ile leu ala arg asn pro arg ser tyr
241/81                               271/91
gtg ggg ggg gag agc gat ccc aac gcg gcc aac ctg gtc cga cac gtt ctc gcc gcc cgc
val gly ala glu ser asp pro asn ala ala asn leu val arg his val leu ala gly arg
301/101                               331/111
ggc gac gtc cgg gtc acc gac gcg gcc gat acc gga tta tcc gac gcc agc gcc gat gtc
gly asp val arg val thr asp ala ala asp thr gly leu ser asp ala ser ala asp val
361/121                               391/131
gtc atc gcc gag gcg atg ctg acc atg caa gcc aac gcg gct aaa cac acg atc gtc gcc
val ile gly glu ala met leu thr met gln gly asn ala ala lys his thr ile val ala
421/141                               451/151
gag cgg gcg cgg gtg ctg agg ccg ggt gcc gcc tac gcg att cac gaa cta gcg ctg gtg
glu ala ala arg val leu arg pro gly gly arg tyr ala ile his glu leu ala leu val
481/161                               511/171
ccg gac gac gtc gca gag cag gtc cgc acc gcc ctg cgg cag tcg ctg gcc cgc gcg ctc
pro asp asp val ala glu gln val arg thr asp leu arg gln ser leu ala arg ala leu
541/181                               571/191
aag gtc aat ggg cgt cgg ctg acc gtt gcc gcc tgg tcg cgc ctc tta gcc gcc cat gga
lys val asn ala arg pro leu thr val ala ala tgg ser thr leu leu ala gly his gly
601/201                               631/211
ctg gtc gtc gac caa gtt gtc acc gct tcc atc tgg tta tta caa cgg cga cgg gty atc
leu val val glu his val val thr ala ser met ala leu leu gln pro arg arg val ile
661/221                               691/231
gct cac gaa gcc ctc ctg ggt gcc ctg gtc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
ala asp glu gly leu leu gly ala leu arg pro ala ala ser leu leu leu leu leu leu
721/241                               751/251
tgg ctt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt
ile arg arg ala val leu leu arg ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt
781/261                               811/271
tgc gtc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc atc
ala val ala ile val ala met lys pro ala val ala leu ala

```

SEQ. N. 421

139/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv2622

```

1/1                                31/11
taa aaa cct gcg gag gtg ggg tct atg gcc aac aaa cgt ggc aat gcc ggg cag cct ctg
OCH lys pro ala glu val gly ser met ala asn lys arg gly asn ala gly gln pro leu
61/21                                91/31
ccc ttg tct gat cga gac gac gac cac atg cag ggg cac tgg ctg ctg gcc cgg ctg ggc
pro leu ser asp arg asp asp asp his met gln gly his trp leu leu ala arg leu gly
121/41                                151/51
aag cgg gtg ctg cgt ccc ggc ggc gtc gaa ctc acc cgg acc ctg ctg gcc cgc gcc gag
lys arg val leu arg pro gly gly val glu leu thr arg thr leu leu ala arg ala glu
181/61                                211/71
gtg acc gac gcc gac gtg ctc gag ctg gca cgg ggc ctg ggc cgc acc gca gcc gaa atc
val thr asp ala asp val leu glu leu ala pro gly leu gly arg thr ala ala glu ile
241/81                                271/91
ttg gcc cgc aac cgg cgg tct tac gtg ggg cgg gag agc gat ccc aac ggc gcc aac ctg
leu ala arg asn pro arg ser tyr val gly ala glu ser asp pro asn ala ala asn leu
301/101                               331/111
gtc cga cac gtt ctc gcc ggc cgc ggc gac gtc cgg gtc acc gac gcc gcc gat acc gga
val arg his val leu ala gly arg gly asp val arg val thr asp ala ala asp thr gly
361/121                               391/131
tta tcc gac gcc agc gcc gat gtc gtc atc ggc gag gcc atg ctg acc atg caa gcc aac
leu ser asp ala ser ala asp val val ile gly glu ala met leu thr met gln gly asn
421/141                               451/151
gcg gct aaa cac acg atc gtc gcc gag gcc gcc cgg gtg ctg agg ccg ggt gcc cgc tac
ala ala lys his thr ile val ala glu ala ala arg val leu arg pro gly gly arg tyr
481/161                               511/171
gcg att cac gaa cta gcg ctg gtg ccg gac gac gtc gca gag cag gtc cgc acc gac ctg
ala ile his glu leu ala leu val pro asp asp val ala glu gln val arg thr asp leu
541/181                               571/191
cgg cag tct ctg gcc cgc gcc ctc aag gtc aat gag cgt ccg ctg acc gtt gcc gaa tgg
arg gln ser leu ala arg ala leu lys val asn ala arg pro leu thr val ala glu trp
601/201                               631/211
tgg cac ctc tta gcc ggc cat gga ctg ctc gtc gaa cac gtt gac acc gct tcc atg gcc
ser his leu leu ala gly his gly leu val val gln his val val thr ala ser met ala
661/221                               691/231
ttg tta caa ccg cga cgg gtg atc gct gac gaa gcc ctc ctg ggt ggg ctg cgg ttc gcc
leu leu gln pro arg arg val ile ala asp glu gly leu leu gly ala leu arg phe ala
721/241                               751/251
gaa aac gtc ctc atc cat ggt gcc gcc atc arg gaa ctc ctc tta atc cac cac aac ttc
gly asn leu leu ile his arg ala ala arg arg val leu leu met arg his thr phe
781/261                               811/271
tgg gtc atc atc gtc gtc ttc arg gcc gtc gcc atc gtc gtc gtc gtc gtc gtc gtc gtc
arg arg his arg gln arg leu thr ala val ala leu val ala ala ggc pro his val asp
841/281
tgg tta
acc gaa

```

ORF d'après Cole et al., 1998

140/185

```

1/1                               31/11
atc gcg cgt gac atc gat gac cag ggt cgg ctg tgt ctg gac gtc gcc ggt cga acg gta
ile ala arg asp ile asp asp gln gly arg leu cys leu asp val gly gly arg thr val
61/21                               91/31
gtt gtt tca gcg gcc gac gtg gtg cat ttg cgt taa ctg gcg cgg agc tgg cgt ccc caa
val val ser ala gly asp val val his leu arg OCH leu ala arg ser trp arg pro gln
121/41                               151/51
aag att aag gtc gcg gcc atg agc tat ccg gag aat gtc ctg gcc gct gcc gag cag gtc
lys ile lys val ala gly met ser tyr pro gln asn val leu ala ala gly glu gln val
181/61                               211/71
gtt ctg cac cgc cat ccg cac tgg aat cgc tta atc tgg ccc gtc gtg gtg ctg gtc ttg
val leu his arg his pro his trp asn arg leu ile trp pro val val val leu val leu
241/81                               271/91
ctg acc ggg ttg gag gag ttc ggg tcc gga ttc gtc aac tgg aca cct tgg cag cag atc
leu thr gly leu ala ala phe gly ser gly phe val asn ser thr pro trp gln gln ile

```

SEQ ID N° 43A

FIGURE 43A

```

1/1                               31/11
tcg cgc gtg aca tgg arg acc agg gtc gcc tgt gtc tgg acc tgg gcg gtc gaa cgg tag
ser arg val thr ser met thr arg val gly cys val trp thr ser ala val glu arg AMB
61/21                               91/31
ttg ttt cag cgg gcg acc tgg tgc att tgc gtt aac tgg cgg gga gct gcc gtc ccc aaa
leu phe gln arg ala thr trp cys ile cys val asn ser ala gly ala gly val pro lys
121/41                               151/51
aga tta agt tca cgg gca tga gct atc cgg aga atg tcc tgg cgg ctg gcg agc agc tgg
arg leu arg ser arg ala OPA ala ile arg arg met ser trp pro leu ala ser arg ser
181/61                               211/71
tgc tgg acc gag atc cgc act gga atc gct taa tgc ggc acc tgc tgc tgc tgc tgc tgc
phe cys thr ala ile arg thr gly ile ala OCH ser gly pro ser trp gln trp ser cys
241/81                               271/91
tga acc att tca cgg gct tca gtc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
OPA pro gly trp arg arg ser gly pro asp pro ser thr arg his leu gly ser arg

```

SEQ ID N° 43B

FIGURE 43B

141/185

```

1/1                               31/11
cgc gcg tga cat cga tga cca ggg tgc gct gtg tct gga cgt cgg cgg tgc aac ggt agt
arg ala OPA his arg OPA pro gly ser ala val ser gly arg arg arg ser asn gly ser
61/21                               91/31
tgt ttc agc ggg cga cgt ggt gca ttt gcg tta act cgc gcg gag ctg gcg tcc cca aaa
cys phe ser gly arg arg gly ala phe ala leu thr arg ala glu leu ala ser pro lys
121/41                               151/51
gat taa ggt cgc ggg cat gag cta tcc gga gaa tgt cct gcc cgc tgg cga gca ggt cgt
asp OCH gly arg gly his glu leu ser gly glu cys pro gly arg trp arg ala gly arg
181/61                               211/71
tct gca cgc cca tcc gca ctg gaa tgc att aat ctg gcc cgt cgt ggt gct ggt cct gct
ser ala pro pro ser ala leu glu ser leu asn leu ala arg arg gly ala gly leu ala
241/81                               271/91
gac cgg gtt gcc gcc gtt cgg gtc cgg att cgt cca ctc gac acc ttg gca gca gat c
asp arg val gly gly val arg val arg ile arg gln leu asp thr leu ala ala asp

```

SEQ ID N° 43C

FIGURE 43C

Séquence codante Rv3278c prédite par Cole et al., 1998 (Nature 393:537-544) et contenant seq43A:

```

1/1                               31/11
atg agc tat ccg gag aat gtc ctg gcc gct ggc ggc gag gtc gtt ctg cac cgc cat ccg
Met ser tyr pro glu asn val leu ala ala gly glu gln val val leu his arg his pro
61/21                               91/31
cac tgg aat cgc tta ata tgg ccc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc
his trp asn arg leu ile trp pro val val val val val val val val val val val val
121/41                               151/51
ttc ggg tcc gga ttc gtc aac tgc acc cct ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
phe gly ser gly phe val asn ser thr pro trp asp gln ile ala lys asn val ile his
181/61                               211/71
gcg gtc ata tgg ggc ata tgg ttg gtc ata ttc ggc tgg gtc ttc gtc tgg cca ttc ctg
ala val ile trp gly ile trp leu val ile val ile trp leu thr leu trp pro phe leu
241/81                               271/91
acc tgg gtc ttc ttc ttc ttc ttc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc gtc
ser trp leu thr thr his phe val val thr thr thr thr thr thr thr thr thr thr thr
301/101                               331/111
ctt aat ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr
361/121                               391/131
tgg ata ttc ggc ggc att ttc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc
arg phe ala ala phe phe arg thr gly thr thr thr thr thr thr thr thr thr thr thr
421/141                               451/151
gac ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
pro leu thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr
481/151                               511/161

```

142/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv3278c

```

1/1                                31/11
taa ctc gcg cgg agc tgg cgt ccc caa aag att aag gtc gcg ggc atg agc tat ccg gag
OCH leu ala arg ser trp arg pro gln lys ile lys val ala gly met ser tyr pro glu
61/21                                91/31
aat gtc ctg gcc gct ggc gag cag gtc gtt ctg cac cgc cat ccg cac tgg aat cgc tta
asn val leu ala ala gly glu gln val val leu his arg his pro his trp asn arg leu
121/41                                151/51
atc tgg ccc gtc gtg gtg ctg gtc ttg ctg acc ggg ttg gcg gcg ttc ggg tcc gga ttc
ile trp pro val val val leu val leu leu thr gly leu ala ala phe gly ser gly phe
181/61                                211/71
gtc aac tgg aca cct tgg cag cag atc gct aag aac gtg att cac gcg gtc atc tgg ggg
val asn ser thr pro trp gln gln ile ala lys asn val ile his ala val ile trp gly
241/81                                271/91
atc tgg ttg gtg atc gtc ggc tgg ctg aag ctg tgg cca ttc ctg agc tgg ctg acc aca
ile trp leu val ile val gly trp leu thr leu trp pro phe leu ser trp leu thr thr
301/101                               331/111
cat ttc gtg gtg acc aac cgg cgg gtg atg ttc cgg cat ggt gtg ctg acc cgc agc ggg
his phe val val thr asn arg arg val met phe arg his gly val leu thr arg ser gly
361/121                               391/131
atc gac ata ccg cta gca cgg atc aac agc gtg gag ttc cgg gac cgg atc ttc gag cgg
ile asp ile pro leu ala arg ile asn ser val glu phe arg asp arg ile phe glu arg
421/141                               451/151
att ttt cgc acc ggg aag ttg att atc gag tcc gcg tca caa gat ccg ctg gag ttc tac
ile phe arg thr gly thr leu ile ile glu ser ala ser gln asp pro leu glu phe tyr
481/161                               511/171
aac att ccc cgc ctg cgg gag gtg cat gag ttg ctg tat caa gag gtt ttc gac acc ctg
asn ile pro arg leu arg glu val his ala leu leu tyr his glu val phe asp thr leu
541/181
ggc tcc gac gag tgg ccc agc tga
gly ser asp glu ser pro ccc GGA

```

SEQ ID NO 43F

FIGURE 43F

143/185

```

1/1                               31/11
gcc aag atg gat gtc tac caa cgc acc gcc gcc gcc tgg cag ccg ctc aag acc ggt atc
ala lys met asp val tyr gln arg thr ala ala gly trp gln pro leu lys thr gly ile
61/21                               91/31
acc acc cat atc ggt tcg gcg gcc atg gcg ccg gaa gcc aag agc gga tat ccg gcc act
thr thr his ile gly ser ala gly met ala pro glu ala lys ser gly tyr pro ala thr
121/41                               151/51
ccg atg ggg gtt tac agc ctg gac tcc gct ttt gcc acc gcg ccg aat ccc ggt gcc ggg
pro met gly val tyr ser leu asp ser ala phe gly thr ala pro asn pro gly gly gly
181/61                               211/71
ttg ccg tat acc caa gtc gga ccc aat cac tgg tgg agt gcc gac gac aat agc ccc acc
leu pro tyr thr gln val gly pro asn his trp trp ser gly asp asp asn ser pro thr
241/81                               271/91
ttt aac tcc atg cag gtc tgt cag aag tcc cag tgc ccg acc agc acg gcc gac agc gag
phe asn ser met gln val cys gln lys ser gln cys pro phe ser thr ala asp ser glu
301/101                               331/111
aac ctg caa atc ccg cag tac aag cat tcg gtc gtg atg gcc gtc aac aag gcc aag gtc
asn leu gln ile pro gln tyr lys his ser val val met gly val asn lys ala lys val
361/121                               391/131
cca gcc aaa gcc tcc gcg ttc ttc ttt cac acc acc gac gcc ggg ccc acc gcc ggt tgt
pro gly lys gly ser ala phe phe phe his thr thr asp gly gly pro thr ala gly cys
421/141
gtg gcg atc
val ala ile

```

SEQ ID N° 44A

FIGURE 44A

```

1/1                               31/11
cca aga tgg atg tct acc aac gca ccg ccg ccg gct ggt agt agt tca aga ccg gta tca
pro arg trp met ser thr asn ala pro pro pro ala gly ser arg ser arg pro val ser
61/21                               91/31
cca cca ata tcg gtt ccg ccg gcc tgg ccg ccg aag cca aga ccg gat atc ccg cca ctc
pro pro ile ser val arg arg ala trp arg arg lys pro arg ala asp ile arg pro leu
121/41                               151/51
cca tgg ggg ttc aaa gcc tgg act ccg ccc ttt gcc acc acc acc acc acc acc acc acc
ala trp gly phe thr ala trp thr pro leu leu ala pro arg arg ile pro val ala gly
181/61                               211/71
tca cca ata tca tgg tgg gac tca atc acc ggt gta tca tca acc acc acc acc acc acc
lys arg ile pro leu ser asp pro ile thr gly gly val ala thr thr ile ala pro pro
241/81                               271/91
tca tca ata tca tgg tgg gac tca atc acc ggt gta tca tca acc acc acc acc acc acc
ala thr pro tyr arg ser val arg ser pro ser ala trp ser ala arg pro thr ala arg
301/101                               331/111
cca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca
thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr
361/121                               391/131
cca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca
thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr
421/141

```

SEQ ID N° 44B

SEQUENCE 44B

144/185

```

1/1                                31/11
caa gat gga tgt cta cca acg cac cgc cgc cgg ctg gca gcc gct caa gac cgg tat cac
gln asp gly cys leu pro thr his arg arg arg leu ala ala ala gln asp arg tyr his
61/21                                91/31
cac cca tat cgg ttc ggc ggg cat ggc gcc gga agc caa gag cgg ata tcc ggc cac tcc
his pro tyr arg phe gly gly his gly ala gly ser gln glu arg ile ser gly his ser
121/41                                151/51
gat ggg ggt tta cag cct gga ctc cgc ttt tgg cac cgc gcc gaa tcc cgg tgg cgg gtt
asp gly gly leu gln pro gly leu arg phe trp his arg ala glu ser arg trp arg val
181/61                                211/71
gcc gta tac cca agt cgg acc caa tca ctg gtg gag tgg cga cga caa tag ccc cac ctt
ala val tyr pro ser arg thr gln ser leu val glu trp arg arg gln AMB pro his leu
241/81                                271/91
taa ctc cat gca ggt ctg tca gaa gtc cca gtc ccc gtt cag cac gcc cga cag cga gaa
OCH leu his ala gly leu ser glu val pro val pro val gln his gly arg gln arg glu
301/101                                331/111
cct gca aac ccc gca gta caa gca ttc ggt cgt gat ggg cgt caa caa ggc caa ggt ccc
pro ala asn pro ala val gln ala phe gly arg asp gly arg gln gln gly gln gly pro
361/121                                391/131
agg caa agg ctc cgc gtt ctt ctt tca cac cac cga cgg cgg gcc cac cgc ggg ttg tgt
arg gln arg leu arg val leu leu ser his his arg arg arg ala his arg gly leu cys
421/141
ggc gat c
gly asp

```

SEQ ID N° 44C

FIGURE 44C



145/135

Séquence codante Rv0309 prédite par Cole et al., 1998 (Nature 393:537-544) et contenant Seq44A:

```

1/1                               31/11
atg agc cga ctc cta gct ttg ctg tgc gct gcg gta tgc acg gcc tgc gtt gct gtg gtt
Met ser arg leu leu ala leu leu cys ala ala val cys thr gly cys val ala val val
61/21                               91/31
ctc gcg cca gtg agc ctg gcc gtc gtc aac ccg tgg ttc gcg aac tgc gtc gcc aat gcc
leu ala pro val ser leu ala val val asn pro trp phe ala asn ser val gly asn ala
121/41                               151/51
act cag gtg gtt tcg gtg gtg gga acc gcc ggt tgc acc gcc aag atg gat gtc tac caa
thr gln val val ser val val gly thr gly gly ser thr ala lys met asp val tyr gln
181/61                               211/71
cgc acc gcc gcc gcc tgg cag ccg ctc aag acc ggt atc acc acc cat atc ggt tgc gcg
arg thr ala ala gly trp gln pro leu lys thr gly ile thr thr his ile gly ser ala
241/81                               271/91
ggc atg gcg ccg gaa gcc aag agc gga tat ccg gcc act tgc atg ggg gtt tac agc ctg
gly met ala pro gln ala lys ser gly tyr pro ala thr pro met gly val tyr ser leu
301/101                               331/111
gac tcc gct ttt gcc acc gcc ccg aat acc ggt gcc ggg ttc ccg tat acc caa gtc gga
asp ser ala phe gly thr ala pro asn pro gly gly gly leu pro tyr thr gln val gly
361/121                               391/131
ccc aat cat tgg tgg agt gcc gac gac aat acc ccc acc ttt aac tgc atg cag gtc tgt
pro asn his trp trp ser gly asp asp asn ser pro thr phe asn ser met gln val cys
421/141                               451/151
cag aag tcc cag tgc ccg ttc agc acg gcc gac agc gag aac ctg caa atc ccg cag tac
gln lys ser gln cys pro phe ser thr ala asp ser gln asn leu gln ile pro gln tyr
481/161                               511/171
aag cat tgc gtc gtg atg gcc gtc aac aag acc acc ttc tgc acc gga aac gcc tcc gcg ttc
lys his ser val val met gly val asn lys ala lys val pro gly lys gly ser ala phe
541/181                               571/191
ttc ttc cag acc acc gac gcc ggg gcc acc ggg ggt ttc tgc acc acc gac gat gcc acc
phe phe his thr thr asp gly gly pro thr ala gly ty val ala ile asp asp ala thr
601/201                               631/211
ctg gtc cag acc atc atc tgc tgc ggt ttc ttc acc ttc ttc acc acc ggt acc taa
leu val gln ile ile arg trp leu arg pro gly ala val ala ala ala ala lys och

```

SEQ ID NO: 441

FIGURE 441

146/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv0309

1/1 31/11  
 tga gcg atg agc cga ctc cta gct ttg ctg tgc gct gcg gta tgc acg ggc tgc gtt gct  
 OPA ala met ser arg leu leu ala leu leu cys ala ala val cys thr gly cys val ala  
 61/21 91/31  
 gtg gtt ctc gcg cca gtg agc ctg gcc gtc gtc aac ccg tgg ttc gcg aac tcg gtc ggc  
 val val leu ala pro val ser leu ala val val asn pro trp phe ala asn ser val gly  
 121/41 151/51  
 aat gcc act cag gtg gtt tcg gtg gtg gga acc gcc ggt tcg acg gcc aag atg gat gtc  
 asn ala thr gln val val ser val val gly thr gly gly ser thr ala lys met asp val  
 181/61 211/71  
 tac caa cgc acc gcc gcc gcc tgg cag ccg ctc aag acc ggt atc acc acc cag acc ggt  
 tyr gln arg thr ala ala gly trp gln pro leu lys thr gly ile thr thr his ile gly  
 241/81 271/91  
 tcg gcg gcc atg gcg ccg gaa gcc aag agc gga tat ccg gcc act ccg atg ggg gtt tac  
 ser ala gly met ala pro glu ala lys ser gly tyr pro ala thr pro met gly val tyr  
 301/101 331/111  
 agc ctg gac tcc gct ttc ggc acc gcg ccg aat ccc ggt gcc ggg ttg ccg tat acc caa  
 ser leu asp ser ala phe gly thr ala pro asn pro gly gly gly leu pro tyr thr gln  
 361/121 391/131  
 gtc gga ccc aat cag tgg tgg agt gcc gac gac aat agc acc acc ttt aac tcc atg cag  
 val gly pro asn his trp trp ser gly asp asp asn ser pro thr phe asn ser met gln  
 421/141 451/151  
 gtc tgt cag aag tcc cag tgc ccg ttc agc atg gcc gac agc gag aac ctg caa atc ccg  
 val cys gln lys ser gln cys pro phe ser thr ala asp ser glu asn leu gln ile pro  
 481/161 511/171  
 cag tac aag cat tcg gtc gtg atg gcc gtc aac acc acc aag gtc acc gcc aaa gcc tcc  
 gln tyr lys his ser val val met gly val asn lys ala lys val pro gly lys gly ser  
 541/181 571/191  
 gcg ttc ttc ttc acc acc acc gac gcc gcc ctc acc acc ggt ggt tct gtg gcc atc gac gat  
 ala phe phe phe his thr thr asp gly gly gln thr ala gly cys val ala ile asp asp  
 601/201 631/211  
 gcc acc ctg ggc cag atc acc ccg tgg ttc acc acc acc acc acc atc gcc atc gcc acc  
 ala thr leu val gln ile ile acc trp leu asp gln gln ala val ile ala ile ala lys  
 661/221  
 tac  
 18

SEQ ID N° 44F

FIGURE 44F

147/185

Fragment cloné en fusion avec phoA

[illegible]

148/185

```

961/321
tca gct cgt cgc tgc ggc gcc gcc ggg ata gaa tgc ccc gcg aac cag tgg tac ggc gca
ser ala arg arg cys gly ala ala gly ile glu ser pro ala asn gln trp tyr gly ala
1021/341
gat tga cct cgt atc atc tga gtt agt tgc ccg cgc aat ggg cat ccg cgt gtt atc ggt
asp opa pro arg ile ile opa val ser cys pro arg asn gly his pro arg val ile gly
1081/361
att acg tga cag tct gtc ggc aag gag gga cgc atg gca ctc tcc gat cat gag cag cgg
ile thr opa gln ser val gly lys glu gly arg met pro leu ser asp his glu gln arg
1141/381
atg ctt gac cag atc gag agc gct ctc tac gcc gaa gat ccc aag ttc gca tgc agt gtc
met leu asp gln ile glu ser ala leu tyr ala glu asp pro lys phe ala ser ser val
1201/401
cgt gcc ggg gcc ttc cgc gca ccg acc ccg ccg cgg cgc ctg cag gcc gcg gag ttc ttc
arg gly gly gly phe arg ala pro thr ala arg arg arg leu gln gly ala ala leu phe
1261/421
atc atc ggt ctg ggg atg ttg gtt tcc gcc gtg gcg ttc aaa gag acc atg atc gga agt
ile ile gly leu gly met leu val ser gly val ala phe lys glu thr met ile gly ser
1321/441
ttc ccg ata ctc agc gtt ttc ggt ttt gtc gtg atg ttc ggt ggt gtg gtg tat gcc atc
phe pro ile leu ser val phe gly phe val val met phe gly gly val val tyr ala ile
1381/461
acc ggt cct cgg ttg tcc gcc agg atg gat cgt gcc gga tgc gct gct ggg gct tgc cgc
thr gly pro arg leu ser gly arg met asp arg gly gly ser ala ala gly ala ser arg
1441/481
cag cgt cgt acc aag ggg gcc ggg gcc tca ttc acc agc cgt atg gaa gat c
gln arg arg thr lys gly ala gly gly ser phe thr ser arg met glu asp
1471/491

```

SEQ ID N° 45ZA (suite)

FIGURE 45ZA (suite)

149/185

fragment seq452A en décalage moins 1 pour la phase de lecture

```

1/1                                31/11
atc tcc ccg gac acc agg tca tcc ggc gag atg gtg atc gag gct cgg acc cgc agg cat
ile ser pro asp thr arg ser ser gly glu met val ile glu ala arg thr arg arg his
61/21                                91/31
ccg gta gcc aga ggc acc agc atc agc aac atc gcg atg gcc agc atg ccg cgc cgt cgg
pro val ala arg gly thr ser ile ser asn ile ala met ala ser met pro arg arg arg
121/41                                151/51
gtc ctt gcc act cgc gat cct tgg gat gac ggt ggg gca tag cta gcg cgc acc agg tca
val leu ala thr arg asp pro trp asp asp gly gly ala AMF leu ala arg thr arg ser
181/61                                211/71
tcg tgc caa acc ggg cat gcc gcg tgg gca agc tgt cgg gcg cgg gtt aga gcg gta gcg
ser cys gln thr gly his ala ala ser ala ser cys arg ala arg val arg ala val ala
241/81                                271/91
tcc cac cca gga tgg cga atg ctc ggg ggt cag cgg cga agt ggt agc cgc gga tga tgt
cys asp pro gly trp arg met leu gly gly his arg arg ser gly ser arg gly OPA cys
301/101                                331/111
cgg tga aac cca acc gcc ggt aca acc gcc agc ccc gat tgt cct cag cgt tgg tct ccg
arg OPA ser pro thr gly gly thr thr ala thr pro asp cys pro his arg trp ser pro
361/121                                391/131
gtg tgg aga gca gga cgt tgt act cgt cgc gac cgg ala gta gtc gcc ggg cca acg cct
val trp arg ala gly arg cys pro arg arg asp arg leu ala val gly gly pro thr pro
421/141                                451/151
ccc cga gcc cac ggc ctt gag cgc ggg gaa ggc tgt gca att cag tca act cga agt agc
pro arg gly his gly leu glu arg gly glu gly cys ala ile gln ser thr arg ser ser
481/161                                511/171
tgg tca tca gtc ggy cga tgg cta gcc gcg cag ccc cga tgg gtt gca agc cca gta cca
trp ser ser val gly arg ser leu gly ala gly ser arg cys val ala ser pro val pro
541/181                                571/191
cct gct gtt gcc acc act ggc cgg gcg ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc
pro ala val ala thr thr gly arg ala pro arg asp ser arg thr pro leu arg ala leu
601/201                                631/211
ggg cgt cga tca gtt cgg cgg cgg cgg cgg cgg cgg cgg cgg cgg cgg cgg cgg cgg cgg
ala arg asp ser val arg arg pro thr ala ala pro thr cys arg arg pro arg pro val
661/221                                691/231
cgg atg arg tta ccc cga cgg cgg cga ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc
arg leu pro leu pro arg arg pro arg pro pro ala ser arg ala ala gly cys ala pro
721/241                                751/251
acc acc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc
ala thr leu gly arg ala leu ser arg y pro thr arg ser ala ser arg arg thr
781/261                                811/271
acc acc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc
ala thr leu gly arg ala leu ser arg y pro thr arg ser ala ser arg arg thr
841/281                                871/291
acc acc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc
ala thr leu gly arg ala leu ser arg y pro thr arg ser ala ser arg arg thr
901/301                                931/311
acc acc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc
ala thr leu gly arg ala leu ser arg y pro thr arg ser ala ser arg arg thr

```

150/185

1021/341	1051/351
att gac ctc gta tca tct gag tta gtt gcc cgc gca atg ggc atc cgc gtg tta tcg gta	
ile asp leu val ser ser glu leu val ala arg ala met gly ile arg val leu ser val	
1081/361	1111/371
tta cgt gac agt ctg tcg gca agg agg gac gca tgc cac tct ccg atc atg agc agc gga	
leu arg asp ser leu ser ala arg arg asp ala cys his ser pro ile met ser ser gly	
1141/381	1171/391
tgc ttg acc aga tcg aga gcg ctc tct acg ccg aag atc cca agt tcg cat cga gtg tcc	
cys leu thr arg ser arg ala leu ser thr pro lys ile pro ser ser his arg val ser	
1201/401	1231/411
gtg gcg ggg gct tcc gcg cac cga ccg cgc ggc ggc ggc tgc agg gcg cgg cgt tgt tca	
val ala gly ala ser ala his arg pro arg gly gly ala cys arg ala arg arg cys ser	
1261/421	1291/431
tca tcg gtc tgg gga tgt tgg ttt ccg gcg tgg cgt tca aag aga cca tga tcg gaa gtt	
ser ser val trp gly cys trp phe pro ala trp arg ser lys arg pro opa ser glu val	
1321/441	1351/451
tcc cga tac tca gcg ttt tcg gtt ttg tcg tca tgt tcg gtg gtg tgg tgt atg cca tca	
ser arg tyr ser ala phe ser val leu ser opa cys ser val val trp cys met pro ser	
1381/461	1411/471
ccg gtc ctc ggt tgt ccg gca gga tgg atc gtg tcg cat cgg cgg ctg ggg ctt cgc gcc	
pro val leu gly cys pro ala gly trp ile val ala asp arg leu leu gly leu arg ala	
1441/481	1471/491
agc gtc gta cca agg ggg ccg ggg gct cat tca cca gcc gta tgg aag atc	
ser val val pro arg gly pro gly ala his ser pro ala val trp lys ile	

SEQ ID N° 452B (suite)

FIGURE 452B (suite)

151/185

fragment seq452A en décalage moins 2 pour la phase de lecture

1/1 31/11  
 tct ccc cgg aca cca ggt cat ccg gcg aga tgg tga tcg agg ctc gga ccc gca ggc atc  
 ser pro arg thr pro gly his pro ala arg trp OPA ser arg leu gly pro ala gly ile  
 61/21 91/31  
 cgg tag cca gag gca cca gca tca gca aca tcg cga tgg cca gca tgc cgc gcc gtc ggg  
 arg AMB pro glu ala pro ala ser ala thr ser arg trp pro ala cys arg ala val gly  
 121/41 151/51  
 tcc ttg cca ctc gcg atc ctt ggg atg acg gtg ggg cat agc tag cgc gca cca ggt cat  
 ser leu pro leu ala ile leu gly met thr val gly his ser AMB arg ala pro gly his  
 181/61 211/71  
 cgt gcc aga ccg gcc atg ccg cgt cgg caa gct gtc ggg cgc ggg tta gag cgg tag cgt  
 arg ala arg pro gly met pro arg arg gln ala val gly arg gly leu glu arg AMB arg  
 241/81 271/91  
 gcg acc cag gat gcc gaa tgc tcg ggg gtc acc gcc gaa gtg gta gcc gcg gat gat gtc  
 ala thr gln asp gly glu cys ser gly val thr gly glu val val ala ala asp asp val  
 301/101 331/111  
 ggt gaa gcc caa ccg gcg gta caa ccg cca cgc ccg att gtc ctc acc gtt ggt ctc cgg  
 gly glu ala gln pro ala val gln pro pro arg pro ile val leu thr val gly leu arg  
 361/121 391/131  
 tgt gga gag cag gac gtt gtc ctc gtc gcg acc gcc tag cag tcg gcg gcc caa cgc ctc  
 cys gly glu gln asp val val leu val ala thr gly AMB gln ser ala gly gln arg leu  
 421/141 451/151  
 ccc gag gcc acg gcc ttg agc gcg ggg aag gat ttg caa ttc agt caa ctc gaa gta gct  
 pro glu ala thr ala leu ser ala gly lys asp val gln phe ser gln leu glu val ala  
 481/161 511/171  
 ggt cat cag tcg gcc gat agc tag gcg cgg aaa gcc gat gcg ttg caa gcc cag tac cac  
 gly his gln ser gly asp arg AMB ala arg lys ala ala ala leu gln ala gln tyr his  
 541/181 571/191  
 ctg ctg ttg cca cca ctg gcc ggg cgc ccc ggg ata gcc gta cgc cac tcc gag cat tgg  
 leu leu leu pro pro leu ala gly arg pro gly ile ala val arg his ser glu his trp  
 601/201 631/211  
 agc gtt gcc cag ttc gcc gcc cga ccg cag cgg cgt ggt ctc ggt ggt ctc gcc ctg ttc  
 arg val ala gln phe gly gly arg arg gln arg arg gly val gly gly leu gly leu pro  
 661/221 691/231  
 gcc tgc cct tac ctc gac gcc ccg gac cgg ctg cca acc ccg ccc ccg gat ctg ctg cag  
 gly cys arg tyr leu asp gly arg asp arg leu pro ala ala pro pro arg val leu gln  
 721/241 751/251  
 cca cat tta gtc cag cca acc ctg gct gcc cct ggt gta acc ctc cag gtc gac ata cag  
 pro his trp tyr ala gln ser leu gly ala pro gly val ala ala arg val asp ctc ttc  
 781/261 811/271  
 ggt cag gtc ctc acc gcc gcc gcc ctg cat atc gtc gtc gtc gtc gtc gtc gtc gtc gtc  
 ala gln gly leu thr gln ala ala leu gln ala ala gly arg gln ala asp ala gln tyr  
 841/281 871/291  
 cca gta cca gcc ctc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc  
 ala gln ala ala val ser ser ser cys asp ala ala met arg ala val ala pro val ser  
 901/301 931/311  
 acg gtc cca gtc ctc ctc ctc ctc ctc ctc ctc ctc ctc ctc ctc ctc ctc ctc ctc ctc  
 arg pro thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

SEQUENCE 452A

PCT/FR98/01813

152/185

961/321 991/331  
 agc tcg tcg ctg cgg cgc cgc cgg gat aga atc gcc cgc gaa cca gtg gta cgg cgc aga  
 ser ser ser leu arg arg arg arg asp arg ile ala arg glu pro val val arg arg arg  
 1021/341 1051/351  
 ttg acc tcg tat cat ctg agt tag ttg ccc cgc caa tgg gca tcc gcg tgt tat cgg tat  
 leu thr ser tyr his leu ser AMB leu pro ala gln trp ala ser ala cys tyr arg tyr  
 1081/361 1111/371  
 tac gtg aca gtc tgt cgg caa gga ggg acg cat gcc act ctc cga tca tga gca gcg gat  
 tyr val thr val cys arg gln gly gly thr his ala thr leu arg ser OPA ala ala asp  
 1141/381 1171/391  
 gct tga cca gat cga gag cgc tct cta cgc cga aga tcc caa gtt cgc atc gag tgt ccg  
 ala OPA pro asp arg glu arg ser leu arg arg arg ser gln val arg ile glu cys pro  
 1201/401 1231/411  
 tgg cgg ggg ctt ccg cgc acc gac cgc cgc ggc ggc cct gca ggg cgc gcc att gtt cat  
 trp arg gln leu pro arg thr asp arg ala ala ala pro ala gln arg gly val val his  
 1261/421 1291/431  
 cat cgg tct ggg gat gtt ggt ttc cgc cgt ggc gtt caa aga gac cat gat cgg aag ttt  
 his arg ser gly asp val gln phe arg arg gly val gln arg asp his asp arg lys phe  
 1321/441 1351/451  
 ccc gat act cag cgt ttt cgg ttt tgt cgt gat gtt cgg tgg tgt ggt gta tgc cat cac  
 pro asp thr gln arg phe arg phe cys arg asp val arg trp cys gly val cys his his  
 1381/461 1411/471  
 cgg tcc tcg gtt gtc ccg cag gat gga tgg tgg cgc atc ggc tgc tgg ggc ttc gcg cca  
 arg ser ser val val arg gln asp gln ser trp arg ile gly cys trp gly phe ala pro  
 1441/481 1471/491  
 gcg tcg tac caa ggg ggc cgg ggg ctc att cac cag ccg tat gga aga tc  
 ala ser tyr gln gly gly arg gly leu ile his gln pro tyr gly arg

SEQ ID N° 45ZC (suite 1)

FIGURE 45ZC (suite 1)

ORF de seq 45CA directement en fusion avec pHA

cag tct gtc gcc aag gag gga ccc atg cca ctc tcc att cat gat tag cgg  
 gln ser val gln lys ile gln arg arg pro leu ser arg ala gln gln arg  
 1141/331 1171/341  
 atg ctt cag tag atc cag arg gat ctt tct ggc gac ctt cca att ttt ttt ttt ttt  
 met leu arg ile ile ile ser ala leu trp ala gln arg pro ile pro ala ser val  
 1201/341 1231/351  
 ctt gtt ggt ggt ttt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt  
 arg gln gln gln phe ala ala pro thr ala arg arg arg leu gln gln ala ala phe  
 1261/341 1291/351  
 atc att att att att att att att att att att att att att att att att att att  
 ile ile gln leu ile met leu val ser gln val ala gln gln thr met ala gln ser  
 1321/341 1351/351  
 ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt ttt  
 met met met met met met met met met met met met met met met met met met met



153/185

Sequence Rv2169c prédite par Cole et al., 1998 (Nature 393:537-544) et contenant Seq45A

```

1/1                                     31/11
atg cca ctc tcc gat cat gag cag cgg atg ctt gac cag atc gag agc gct ctc tac gcc
Met pro leu ser asp his glu gln arg met leu asp gln ile glu ser ala leu tyr ala
61/21                                     91/31
gaa gat ccc aag ttc gca tgc agt gtc cgt ggc ggg ggc ttc cgc gca ccg acc gcg cgg
glu asp pro lys phe ala ser ser val arg gly gly gly phe arg ala pro thr ala arg
121/41                                     151/51
cgg cgc ctg cag ggc gcg gcg ttg ttc atc atc ggt ctg ggg atg ttg gtt tcc ggc gtg
arg arg leu gln gly ala ala leu phe ile ile gly leu gly met leu val ser gly val
181/61                                     211/71
gcg ttc aaa gag acc atg atc gga agt ttc ccg ata ccc agc gtt ttc ggt ttt gtc gtg
ala phe lys glu thr met ile gly ser phe pro ile leu ser val phe gly phe val val
241/81                                     271/91
atg ttc ggt ggt gtg gtg tat gcc atc acc ggt cct cgg ttg tcc ggc agg atg gat cgt
met phe gly gly val val tyr ala ile thr gly pro arg leu ser gly arg met asp arg
301/101                                    331/111
ggc gga tgc gct gct ggg gct tgc cgc cag cgt cgt acc aag agg gcc ggg ggc tca ttc
gly gly ser ala ala gly ala ser arg gln arg arg thr lys gly ala gly gly ser phe
361/121                                    391/131
acc agc cgt atg gaa gat cgg ttc cgg cgc cgc ttc acc gag taa
thr ser arg met glu asp arg phe arg arg phe arg glu LCH

```

SEQ ID N° 45D

FIGURE 45D

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv2169c

```

1/1                                     31/11
tga cag cct gtc ggc aag gag gga cga atc acc cct cgc gat cat gaa cag cgg atg ctt
OPA gln ser val gly lys gln gly ala met pro leu ser asp gln glu gln arg met leu
61/21                                     91/31
gac cag atc gag agc gct ctc tat ggc cag gat acc ggt ttc ggt ttt gtc gtg
asp gln ile glu ser ala leu tyr ala ala arg pro thr ala ser val arg ala
121/41                                     151/51
cag ggc ctc ccc gcc acc acc ggt cgg acc acc ctc acc acc acc ttt ttc atc atc
gly gly phe arg ala pro thr ala arg arg arg arg arg arg arg arg arg arg arg
181/61                                     211/71
atg ctc acc acc ttc gtt ttc ggt ttc ggt ttc ggt ttc ggt ttc ggt ttc ggt ttc ggt
gly leu arg met leu val leu gly val ala phe gly arg thr met ala glu ser phe pro
241/81                                     271/91
gaa ggc agc ggt ttc ggt ttc ggt ttc ggt ttc ggt ttc ggt ttc ggt ttc ggt ttc ggt
arg leu arg val phe gly phe val val met phe arg arg arg arg arg arg arg arg arg
301/101                                    331/111
acc agc cgt atg gaa gat cgg ttc cgg cgc cgc ttc acc gag taa
thr ser arg met glu asp arg phe arg arg phe arg glu LCH

```

154/185

```

1/1
cag ccg cgc cgc atc gac cag ggc ctc acg ccc ggt cac ttc tcc gcg ttc ctc aac aat
gln pro arg arg ile asp gln gly leu thr pro gly his phe ser ala phe leu asn asn
61/21
tcc ggt gaa cat cgc acc agg tta ggc agc aat ccc gcg gac ccg cac ccc act cgc cga
ser gly glu his arg thr arg leu gly ser asn pro ala asp pro his pro thr arg arg
121/41
ccg gcc aac tca cag aca ccc tct acg atg cag ggt atg cgg acc ccc aga cgc cac tgc
pro ala asn ser gln thr pro ser thr met gln gly met arg thr pro arg arg his cys
181/61
cgt cgc atc gcc gtc ctc gcc gcc gtt agc atc gcc gcc act gtc gtt gcc gcc tgc tcg
arg arg ile ala val leu ala ala val ser ile ala ala thr val val ala gly cys ser
241/81
tcg gcc tcg aag cca agc gcc gga cca ctt cgg gac gcg aag ccg ctg gtc gag gag gcc
ser gly ser lys pro ser gly gly pro leu pro asp ala lys pro leu val glu glu ala
301/101
acc gcg cag acc aag gct ctc aag agc gag gtc atg atg ctg acg gtc aac gcc aag atc
thr ala gln thr lys ala leu lys ser ala his met val leu thr val asn gly lys ile
331/111

```

SEQ ID N° 46A

FIGURE 46A

[illegible]

155/185

1/1  
gcc gcg ccg cat cga cca ggg cct cac gcc 31/11  
ala ala pro his arg pro gly pro his ala arg ser leu leu arg val pro gln gln phe  
61/21  
cgg tga aca tcg cac cag gtt agg cag caa tcc cgc gga ccc gca ccc cac tcg ccg acc  
arg opa thr ser his gln val arg gln gln ser arg gly pro ala pro his ser pro thr  
121/41  
ggc caa ctc aca gac acc ctc tac gat gca ggg tat ggg gac ccc cag acg cca ctg ccg  
gly gln leu thr asp thr leu tyr asp ala gly tyr ala asp pro gln thr pro leu pro  
181/61  
tcg cat cgc cgt cct cgc cgc cgt tag cat cgc cgc cac tgt cgt tgc ccg ctg ctc gtc  
ser his arg arg pro arg arg arg AMB his arg arg his cys arg cys arg leu leu val  
241/81  
ggg ctc gaa gcc aag ccg ccg acc act tcc gga cgc gaa ggc ggc ggt cga gga ggc cag  
gly leu glu ala lys arg arg thr thr ser gly arg glu ala ala gly arg gly gly his  
301/101  
cgg gca gac caa ggc tct caa gag cgc gca cat ggt gct gac ggt caa cgc caa gat c  
arg ala asp gln gly ser gln glu arg ala his gly ala asp gly gln arg gln asp  
331/111

SEQ ID N° 46C

FIGURE 46C

156/185

Séquence codante Rv1411c prédite par Cole et al., 1998 (Nature 393: 537-544) et contenant seq46A:

```

1/1                               31/11
atg cgg acc ccc aga cgc cac tgc cgt cgc atc gcc gtc ctc gcc gcc gtt agc atc gcc
Met arg thr pro arg arg his cys arg arg ile ala val leu ala ala val ser ile ala
61/21                               91/31
gcc act ctc gtt gcc ggc tgc tcg tcg gcc tcg aag cca agc ggc gga cca ctc cgg gac
ala thr val val ala gly cys ser ser gly ser lys pro ser gly gly pro leu pro asp
121/41                               151/51
ggc aag cgg ctg gtc gag gag gcc acc gcc cag acc aag gct ctc aag agc gcc cac atg
ala lys pro leu val glu glu ala thr ala gln thr lys ala leu lys ser ala his met
181/61                               211/71
gtg ctg acg gtc aac ggc aag atc cgg gga ctg tct ctg aag acg ctg agc gcc gat ctc
val leu thr val asn gly lys ile pro gly leu ser leu lys thr leu ser gly asp leu
241/81                               271/91
acc acc aac ccc acc gcc ggc acg gga aac ctc aag ctc aag ctg ggt ggg tct gat atc
thr thr asn pro thr ala ala thr gly asn val lys leu thr leu gly gly ser asp ile
301/101                               331/111
gat gcc gac ttc gtg gtg ttc gac ggg atc ctg tac gcc acc ctg acg ccc aac cag tgg
asp ala asp phe val val phe asp gly ile leu tyr ala thr leu thr pro asn gln trp
361/121                               391/131
agc gat ttc ggt ccc gcc gcc gac atc tac gac ccc gcc cag gtg ctg aat cgg gat acc
ser asp phe gly pro ala ala asp ile tyr asp pro ala gln val leu asn pro asp thr
421/141                               451/151
ggc ctg gcc aac gtg ctg gcg aat ttc gcc gac gaa aaa gcc gaa ggg cgg gat acc atc
gly leu ala asn val leu ala asn phe ala asp ala lys ala glu gly arg asp thr ile
481/161                               511/171
aac ggc cag aac acc atc cgc atc agc ggg aag gta tcg cca cag gcc gtg aac cag ata
asn gly gln asn thr ile arg ile ser gly lys val ser ala gln ala val asn gln ile
541/181                               571/191
ggc cgg cgg ttc aac gcc acg cag cgg ctg ccc ggg acc atc tgg att cag gag acc gcc
ala pro pro phe asn ala thr gln pro val pro ala thr val tip ile gln gln thr gly
601/201                               631/211
gat cat cca ctg gca cag gcc cag tgg gta cgc gta tgg ggc atc tcc acc cag atc acc
asp his gln leu ala gln ala gln leu asp arg gly ccc gly acc ser val ala met thr
661/221                               691/231
tct tgg aac tgg acc gta acc ctc acc ctc acc acc acc acc acc acc acc acc acc acc
leu asp lys thr gly ala lys val ala val thr lys pro pro val pro pro

```

SEQ ID N° 46E

FIGURE 46D

157/185

ORF d'après par Cole et al., 1998 (Nature 393: 537-544):  
et contenant la séquence codante Rv1411c:

```

1/1                               31/11
tag ctc acc cag gtt gga ccg gtt cag tgt ctc ggc cat cac gtc ggc ggt gaa ttg gcc
AMB leu thr gln val gly pro val gln cys leu gly his his val gly gly glu leu ala
61/21                               91/31
gtc ggg caa tac atc gac gac cgt cag aca cac gcc gtt gac agc gat cga gtc gcc gtg
val gly gln tyr ile asp asp arg gln thr his ala val asp ser asp arg val ala val
121/41                               151/51
gcc ggc gtc ggc ggt aac cat cgg acc gcc gat ggt cag ccg cgc cgc atc gac cag gcc
ala gly val gly gly asn his arg thr ala asp gly gln pro arg arg ile asp gln gly
181/61                               211/71
ctc acg ccc ggt cac ttc tcc gcg ttc ctc aac aat tcc ggt gaa cat cgc acc agg tta
leu thr pro gly his phe ser ala phe leu asn arg ser gly glu his arg thr arg leu
241/81                               271/91
ggc agc aat ccc gcg gac ccg cac ccc act cgc cga ccg gcc aac tca cag aca ccc tct
gly ser asn pro ala asp pro his pro thr arg arg pro ala asn ser gln thr pro ser
301/101                               331/111
acg atg cag ggt atg ccg acc ccc aga cgc cac ttc tcc ggt gaa cat cgc acc agg tta
thr met gln gly met arg thr pro arg arg his cys arg arg ile ala val leu ala ala
361/121                               391/131
gtt agc atc gcc gcc act gln gtt gcc ggc tgc tgc tgc ggc tgc aag cca agc gcc gga
val ser ile ala ala thr val val ala gly cys ser ser gly ser lys pro ser gly gly
421/141                               451/151
cca ctt ccg gac gcg aag ccg ctg gtc gag gag gcc acc gcg cag acc aag gct ctc aag
pro leu pro asp ala lys pro leu val glu glu ala thr ala gln thr lys ala leu lys
481/161                               511/171
agc gcg cac atg gtg ctg acg gtc aac gcc aag atc ccg gga ctg tct ctg aag acg ctg
ser ala his met val leu thr val asn gly lys ile pro gly leu ser leu lys thr leu
541/181                               571/191
agc gcc gat ctc acc acc aac ccc acc gcc ggt gtc ttc aat ttc aag ctc aag ctg ggt
ser gly asp leu thr thr asn pro thr ala ala ttc gly asp val lys leu thr leu gly
601/201                               631/211
ggg tct gat atc gat gcc gac ttc ctg gtc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
gly ser asp ile asp ala asp phe val val phe arg gly ile leu tyr ala thr leu thr
661/221                               691/231
ccc aac cag tgg agc gat ttc ggt ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc ccc
pro asn gln tip ser asp phe gly pro ala ala ccc ccc ccc ccc ccc ccc ccc ccc ccc
721/241                               751/251
aat ccc cat ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
asn pro asp thr ile leu ala asn val leu ala ccc ccc ccc ccc ccc ccc ccc ccc ccc
781/261                               811/271
ggg gtc aac ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
arg asp thr ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
841/281                               871/291
ggg aac ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc ttc
val asn gln ile ala pro pro phe asn ala thr thr thr thr thr thr thr thr thr thr
901/301                               931/311
tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
ile thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr
961/321                               991/331
tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
ile thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

```

TABLE 1: Nucleotide sequence of Rv1411c

TABLE 2: Amino acid sequence of Rv1411c

158/185

```

1/1                               31/11
gag ctg gtc aac ggc gcc ggc atc gac gac gcc gcc gtc gtg acc tgc cgg ccg gac agc
glu leu val asn gly ala gly ile asp asp ala ala val val thr cys arg pro asp ser
61/21                               91/31
ctg gcc gat gcc cag cag atg gtc gag gcg gca ctg ggc cga tat ggc cgt ttg gac qga
leu ala asp ala gln gln met val glu ala ala leu gly arg tyr gly arg leu asp gly
121/41                               151/51
gtg ttg gtg gcc tcg ggc agc aac cat gtg gcg gcc att acc gag atg gcc gtc gag gac
val leu val ala ser gly ser asn his val ala pro ile thr glu met ala val glu asp
181/61                               211/71
ttc gac gct gtg atg gac gcg aac gtg cgg ggt gcc tgg ctg gtg tgt cgg gcg gcc gga
phe asp ala val met asp ala asn val arg gly ala tip leu val cys arg ala ala gly
241/81                               271/91
cgg gtg ctg ctg gag cag ggt cag gcc ggc agc gtg gtg ctg gtg tcg tcc gtt cgc gcc
arg val leu leu glu gln gly gln gly gly ser val val leu val ser ser val arg gly
301/101                               331/111
ggg ttg gcc aat gcc gcc ggt tac agc gcg tac tgc cgg tgg aag gcg gcc acc gat c
gly leu gly asn ala ala gly tyr ser ala tyr cys pro ser lys ala gly thr asp

```

SEQ ID N° 47A

FIGURE 47A

```

1/1                               31/11
agg tgg tca aag gag cgg gcc tcc aag acc cgg cgg tgg tga cct gcc gcc cgg acc gcc
ser trp ser thr ala pro ala ser thr thr pro pro ser CPA pro ala gly arg thr ala
61/21                               91/31
tgg cgg atg ccc agt aga tgg tcc agc cgg acc ttc acc gat atg gcc gtt tgg aag ggt
trp pro met pro ser arg trp ser arg arg trp ala asp met ala val trp thr gly
121/41                               151/51
tgc tgg tgg tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
cys trp trp pro ala ala ala thr met tip agc gcc gcc pro ala trp pro ser arg thr
181/61                               211/71
tgc tgg tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
ser thr leu CPA tip thr ala thr cys gly val trp gly tip ty val gly arg thr ala
241/81                               271/91
tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
gly cys cys ser ser val ala ala ala tip ty tip cys ala pro phe ala ala
301/101                               331/111
tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
trp thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr thr

```

159/185

```

1/1                               31/11
gct ggt caa cgg cgc cgg cat cga cga cgc cgc cgt cgt gac ctg ccg gcc gga cag cct
ala gly gln arg arg arg his arg arg arg arg arg arg asp leu pro ala gly gln pro
61/21                               91/31
ggc cga tgc cca gca gat ggt cga ggc ggc act ggg ccg ata tgg ccg ttt gga cgg agt
gly arg cys pro ala asp gly arg gly gly thr gly pro ile trp pro phe gly arg ser
121/41                               151/51
gtt ggt ggc ctg ggg cag caa cca tgt ggc gcc cat tac cga gat ggc cgt cga gga ctt
val gly gly leu gly gln gln pro cys gly ala his tyr arg asp gly arg arg gly leu
181/61                               211/71
cga cgc tgt gat gga cgc qaa cgt gcg ggg tac ctg act ggt gtg tgg ggc ggc cgg acg
arg arg cys asp gly arg glu arg ala gly cys leu ala gly val ser gly gly arg thr
241/81                               271/91
ggt gct gct cga gca ggg tca ggg cgg cag cgt ggt gct ggt gtc gtc cgt tgg cgg cgg
gly ala ala arg ala gly ser gly arg gln arg gly ala gly val val arg ser arg arg
301/101                               331/111
gtt ggg caa tgc cgc cgg tta cag cgc gta ctg ccg gtc gaa ggc ggg cac cga tc
val gly gln cys arg arg leu gln arg val leu pro val glu gly gly his arg

```

SEQ ID N° 47C

FIGURE 47C

160/185

Séquence codante Rv1714 prédite par Cole et al., 1998 (Nature 393: 537-544) et contenant seq 47A:

```

1/1                               31/11
gtg gag gaa atg gcg ctg gct cag cag gtg ccg aac ctg ggt ctg gcg cgc ttc agc gtg
val glu glu met ala leu ala gln gln val pro asn leu gly leu ala arg phe ser val
61/21                               91/31
cag gac aag tcg atc ctg atc acc gcc gcg acc ggt tcg ttg gcc cga gtt gcc gcc cgg
gln asp lys ser ile leu ile thr gly ala thr gly ser leu gly arg val ala ala arg
121/41                               151/51
gcg ctg gcc gac gcg gga gcg cgg ctg aca ctg gcc gcc gcc aac tcg gcc ggt ctg gcc
ala leu ala asp ala gly ala arg leu thr leu ala gly gln asn ser ala gly leu ala
181/61                               211/71
gag ctg gtc aac gcc gcc gcc atc gac gac gcc gcc gtc gtg acc tgc cgg ccg gac agc
glu leu val asn gly ala gly ile asp asp ala ala val val thr cys arg pro asp ser
241/81                               271/91
ctg gcc gat gcc cag cag atg gtc gag gcg gca ctg gcc cga tat gcc cgt ttg gac gga
leu ala asp ala gln gln met val glu ala ala leu gly arg tyr gly arg leu asp gly
301/101                               331/111
gtg ttg gtg gcc tcg gcc agc aac cat gtg gcg gcc att acc gag atg gcc gtc cag gac
val leu val ala ser gly ser asn his val ala pro ile thr glu met ala val glu asp
361/121                               391/131
ttc gac gct gtg atg gac gcg aac gtg cgg ggt gcc tgg ctg gtg tgt cgg gcg gcc gga
phe asp ala val met asp ala asn val arg gly ala trp leu val cys arg ala ala gly
421/141                               451/151
cgg gtg ctg ctc gag cag ggt cag gcc gcc agc gtg gtg ctg gtg tcg tcc gtt cgc gcc
arg val leu leu glu gln gly gln gln gly ser val val leu val ser ser val arg gly
481/161                               511/171
ggg ttg gcc aat gcc gcc ggt tac agc gcg tac tgc ccg tcg aag gcc gcc acc gat ctg
gly leu gly asn ala ala gly tyr ser ala tyr cys pro ser lys ala gly thr asp leu
541/181                               571/191
ttg gcc aag aca ttg ccg gcc aca ttg gcc ggt cag gcc acc gcc gtg acc gcc ctg gcc
leu ala lys thr leu ala ala glu trp gly gln his gly ile arg val asn ala leu ala
601/201                               631/211
ccg acc gtg ttc cgg tcg gcc gtg acc gag tgg atg ttc acc gac tat ccg aag gcc ccg
pro thr val phe arg ser ala val thr glu trp met phe thr asp asp pro lys gly arg
661/221                               691/231
gcc acc cga gac gcc atg gcc gcc ccg acc acc tcg gcc acc tcg gcc acc gcc acc gcc
ala thr arg ala ala met leu ala ala ile pro leu acc asp pro ala ala pro ala asp
721/241                               751/251
ttc gtc gcc gcc ctg atc tat gcc pro acc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
phe val arg ala leu ala thr val leu leu ser asp ala gcc ser pro tyr thr gly ala val
781/261                               811/271
atg tat gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc
met tyr leu asp gly gln tyr thr ala cys gcc

```



161/185

ORF d'après Cole et al., 1998 (Nature 393: 537-544) et contenant la séquence codante Rv1714:

24/1 54/11  
tag gtg gag gaa atg gcg ctg gct cag cag gtg ccg aac ctg ggt ctg gcg cgc ttc agc  
AMB val glu glu met ala leu ala gln gln val pro asn leu gly leu ala arg phe ser  
84/21 114/31  
qtg cag gac aag tcg atc ctg atc acc gcc gcc acc ggt tcg ttg gcc cga gtt gcc gcc  
val gln asp lys ser ile leu ile thr gly ala thr gly ser leu gly arg val ala ala  
144/41 174/51  
cgg gcg ctg gcc gac gcg gga gcg cgg ctg acc ctg gcc gcc gcc aac tcg gcc ggt ctg  
arg ala leu ala asp ala gly ala arg leu thr leu ala gly gly asn ser ala gly leu  
204/61 234/71  
gcc gag ctg gtc aac gcc gcc gcc atc gac gcc gcc gcc gtc gtg acc tgc cgg ccg gac  
ala glu leu val asn gly ala gly ile asp asp ala ala val val thr cys arg pro asp  
264/81 294/91  
ajc ctg gcc gat gcc cag cag atg gtc gag gcc gcc gcc gcc gcc tat gcc cgt ttg gac  
ser leu ala asp ala gln gln met val glu ala ala leu gly arg tyr gly arg leu asp  
324/101 354/111  
gja gtg ttg gtg gcc tcg gcc agc aac cat gtg gcc gcc att acc gag atg gcc gtc gag  
gly val leu val ala ser gly ser asn his val ala pro ile thr glu met ala val glu  
384/121 414/131  
gac ttc gac gct gtg atg gac gcg aac gtg cgg ggt gcc tcg ctg gtg tgt cgg gcc gcc  
asp phe asp ala val met asp ala asn val arg gly ala trp leu val cys arg ala ala  
444/141 474/151  
gja cgg gtg ctg ctc gag cag ggt cag gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
gly arg val leu leu glu gln gly gln gly gly ser val val leu val ser ser val arg  
504/161 534/171  
gic ggg ttg gcc aat gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
gly gly leu gly asn ala ala gly tyr ser ala tyr cys pro ser lys ala gly thr asp  
564/181 594/191  
ctg ttg gcc aag acc ttg gcg gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
leu leu ala lys thr leu ala ala glu trp gly gly his gly ile arg val asn ala leu  
624/201 654/211  
ggy gcc acc gtg ttg gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
ala pro thr val phe arg ser ala val thr ala thr met pro thr ala asp pro lys gly  
684/221 714/231  
gic gcc acc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
arg ala thr arg ala ala met val ala arg ala thr leu ala arg pro ala ala pro ala  
744/241 774/251  
gic gcc acc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
arg phe val gly ala leu ile tyr leu leu gcc gcc ala ser ser pro thr thr gly ala  
804/261 834/271  
gic gcc acc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
val met thr leu ala gly gly tyr thr ala arg arg

162/185

```

1/1                               31/11
agg ctc atg agc aag acg gtt ctc atc ctt ggc gcg ggt gtc ggc ggc ctg acc acc gcc
arg leu met ser lys thr val leu ile leu gly ala gly val gly gly leu thr thr ala
61/21                             91/31
gac acc ctc cgt caa ctg cta cca cct gag gat c
asp thr leu arg gln leu leu pro pro glu asp

```

SEQ ID N° 48A

FIGURE 48A

```

1/1                               31/11
ggc tca tga gca aga cgg ttc tca tcc ttg gag cgg gtc tgg ggc gcc tga cca cgg cgg
gly ser opa ala arg arg phe ser ser leu ala arg val ser ala ala opa pro pro pro
61/21                             91/31
aca ccc tcc gtc aac tgc tac cac ctg agg atc
thr pro ser val asn cys tyr his leu arg ile

```

SEQ ID N° 48B

FIGURE 48B

```

1/1                               31/11
gag tca tca gca aga cgg ttc tca tcc ttg gag cgg gtc tgg ggc gcc tga cca cgg cgg
gly ser opa ala arg arg phe ser ser leu ala arg val ser ala ala opa pro pro pro
61/21                             91/31
aca ccc tcc gtc aac tgc tac cac ctg agg atc
thr pro ser val asn cys tyr his leu arg ile

```

SEQ ID N° 48C

FIGURE 48C

163/185

[illegible]

164/185

ORF d'après Cole et al., 1998 (Nature 393: 537-544) et contenant la séquence codante Rv0331:

```

1/1                               31/11
tga aca ccc gcg ccg acg cgg cga caa tgc cgg aaa acc ggt ccg cgg gaa tgc tgc ggg
OPA thr pro ala pro thr arg arg gln ser arg lys thr gly pro arg glu cys cys gly
61/21                               91/31
cca tgg gcc gat aat agt ttg act gac tgc gtc agt cac ccc aag acc ttg cgc aag act
pro trp ala asp asn ser leu thr asp ser val ser his pro lys thr leu arg lys thr
121/41                               151/51
gcg gcg gaa tct aat att cca aag ata tat gga act cga tgc gaa gga atc agg ctc atg
ala ala glu ser asn ale pro lys ile tyr gly thr arg cys glu gly ile arg leu met
181/61                               211/71
agc aag acg gtt ctc atc att gcc gcg ggt gtc gcc gcc ctg acc acc gcc gac acc ctc
ser lys thr val leu ile leu gly ala gly val gly gly leu thr thr ala asp thr leu
241/81                               271/91
cgt caa ctg cta cca cct gag gat cga atc ata ttg gtg gac agg agc ttt gac ggg acg
arg gln leu leu pro pro glu asp arg ile ile leu val asp arg ser phe asp gly thr
301/101                               331/111
ctg gcc ttg tgc ttg cta tgg gtg ttg cgg gcc tgg cgg cgg cct gac gac gtc cgc gtc
leu gly leu ser leu leu trp val leu arg gly trp arg arg pro asp asp val arg val
361/121                               391/131
cgc ccc acc gcg gcg tgc ctg ccc ggt gtg gaa atg gtc act gca acc gtc gcc cac att
arg pro thr ala ala ser leu pro gly val glu met val thr ala thr val ala his ile
421/141                               451/151
gac atc gcg gcc cag gta gtg cac acc gac aac agc gtc atc gcc tat gac gcg ttg gtg
asp ile ala ala gln val val his thr asp asn ser val ile gly tyr asp ala leu val
481/161                               511/171
atc gca tta ggt gcg gcg ctg aac acc gac gcc gtt ccc gga ctg tgc gac gcg ctc gac
ile ala leu gly ala ala leu asn thr asp ala val pro gly leu ser asp ala leu asp
541/181                               571/191
gcc gac gtc gcg gcc cag ttc tac acc ctg gac gcc ggt gat gac ctg cgt gcg aag gtc
ala asp val ala gly gln phe tyr thr leu asp gly ala ala ala leu arg ala lys val
601/201                               631/211
gag gcg ctc gag cat gcc cgg atc gat gtg gct atc ccc ggg gta cgg ttc aaa tgc cca
glu ala leu ala his gly asp ile ala val ala ile ala ala gly val pro phe lys cys pro
661/221                               691/231
acc gaa cgt taa aac aac aac ttc cta atc aac aac cgt ggt gac cgt taa gcc acc
ala ala pro pro ala ala phe leu ile ala ala ala leu ala asp arg tyr ala thr
721/241                               751/251
gta acc gta aat atc aac aac ttc aat cgt gac gta cgt gta cgt gta cgt gta cgt gta
acc thr val ala ala ala phe thr pro asp gta leu pro met pro val ala gly pro
781/261                               811/271
gac gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta
ala val gly ala ala leu val ser met leu lys asp ala gly val gly phe his pro arg
841/281                               871/291
gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta
gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta gta

```

165/185

```

901/301
gaa ccg ttc gat ctg ctt gcc gtg gtc ccc ccg cac gtg ccc tcc gcc gcg gcg cgg tca
glu pro phe asp leu leu ala val val pro pro his val pro ser ala ala ala arg ser
961/321
gcg ggt ctc agc gaa tcc ggg tgg ata ccc gtg gac ccg cgc acc ctg tcc act agc gcc
ala gly leu ser glu ser gly trp ile pro val asp pro arg thr leu ser thr ser ala
1021/341
gac aac gtg tgg gcc atc gcc gat gcg acc gtg ctg acg ctg ccg aat gcc aaa cgg ctg
asp asn val trp ala ile gly asp ala thr val leu thr leu pro asn gly lys pro leu
1081/361
ccc aag gct gcc gtg ttc gcc gaa gcc cag gcc gaa gtt gtc gcc cac gcc gtc gcc cgc
pro lys ala ala val phe ala glu ala gln ala ala val val ala his gly val ala arg
1141/381
cat ctc ggt tac gac gta gct gag cgc cac ttc acc gcc aag gcc gcc tgc tac gtc gag
his leu gly tyr asp val ala glu arg his phe thr gly thr gly ala cys tyr val glu
1201/401
acc ggt gat cac cag gca gcc aag gcc gac gcc gat ttc ttc gct ccg tgg gcg ccc tgg
thr gly asp his gln ala ala lys gly asp gly asp phe phe ala pro ser ala pro ser
1261/421
gtg acg ctg tac ccg ccg tgg cgg gag ttt cac gag gag aag gtc gca caa gaa ctg gcc
val thr leu tyr pro pro ser arg glu phe his glu glu lys val ala gln glu leu ala
1321/441
tgg ctg acc cgc tgg aag acg tga
trp leu thr arg trp lys thr OPA

```

SEQ ID N° 48F (suite)

FIGURE 48F (suite)

166/185

Fragment amplifié par PCR d'après les similarités de séquences avec une sérine protéase de la famille htrA de E. coli (création du site BamHI à l'extrémité 5' et du site SnaBI à l'extrémité 3') et sous-cloné dans le vecteur pJVED:

```

1/1
cca tct aca ccg ctc aac agc cgg gcc aga cgc tgc cgg tcg gtg ctg ccg aga agg cgg
pro ser thr pro leu asn ser arg ala arg arg cys arg ser val leu pro arg arg arg
61/21
tga tcc gtg gcg agt tgt tca tgt cgc gcc gca cca ccg ccg acc aac ggg tgc atg cca
OPA ser val ala ser cys ser cys arg gly ala pro pro pro thr asn gly cys leu pro
121/41
tcc gtc tga cca acg qta gtt cgc tgc tga tct cca aaa gtc tca agc cca ccg aag cag
ser val OPA pro thr val val arg cys OPA ser pro lvs val ser ser pro pro lys gln
181/61
tca tga aca agc tgc gtt ggg tgc tat tga tgc tgg gtg gga tgc ggg tgg cgg tgc ccg
ser OPA thr ser cys val gly cys tyr OPA ser trp val gly ser gly trp arg ser pro
241/81
cgg tgg ccg ggg gga tgg tca ccc ggg ccg ggc tga ggc tgg tgg gcc gcc tca ccg aag
arg trp pro gly gly trp ser pro gly pro gly OPA gly arg trp ala ala ser pro lys
301/101
cgg ccg agc ggg tgg ccg gaa ccg acc tgc gcc cca tcc ccg tct tgc gca gcc acc
arg pro ser gly trp arg glu pro thr thr cys gly pro ser pro ser ser ala ala thr
361/121
aat tgg cca gcc tga cag agg cat tca att taa tgc tgc ggg cgc tgg ccg agt ccc ggg
asn trp pro gly OPA gln arg his ser ile OCH cys cys gly arg trp pro ser his gly
421/141
aac gcc agc caa gcc tgg tta ccg acc ccg gac atg aat tgc gta ccc cgc taa cgt cgc
asn gly arg gln gly trp leu pro thr pro asp met asn cys val pro arg OCH arg arg
481/161
tgc gca cca atg tgc aac tct tga tgg cct gga tgg ccc cgc ggg ctc cgc gcc taa cca
cys ala pro met ser asn ser OPA trp pro arg trp pro arg gly leu arg gly tyr pro
541/181
agc agg aga tgg tgc acc tgc atg ccc atg tgc tgg ctc aaa tgc agg aat tat cca cag
ser arg arg trp ser thr cys val pro met cys trp leu lys ser arg asn cys pro his
601/201
tgg tag gcc att tgg tgg acc tct ccc gac gac acc ccc gag aag tgg tgc acc arg cgg
trp AMB ala ile tip tip thr cys pro ala ala thr pro ala lys trp cys thr ser arg
661/221
tgc gca tgc ctc acc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
ser thr trp leu thr ser ser thr ala ala tip ser gly ser gly gly gly ala thr ala
721/241
tgc tct tgc aat tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
ser pro ser thr ser thr OPA leu gly gly arg phe met ala ala pro leu asp cys arg
781/261
gga tgg cca tgc acc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
gly trp arg leu thr OPA tip thr thr thr arg ser gly ala ala ala ala ala thr trp
841/281
tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc tgc
val ser ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala ala

```

167/185

901/301  
 cgg gca ttc ccg tgc agg agc gcc gtc tgg tgt ttg aac ggt ttt acc ggt cgg cat cgg  
 arg ala phe pro cys arg ser ala val trp cys leu asn gly phe thr gly arg his arg  
 961/321  
 cac ggg cgt tgc cgg gtt cgg gcc tgc ggt tgg cga tgc tca aac agg tgg tgc tca acc  
 his gly arg cys arg val arg ala ser gly trp arg ser ser asn arg trp cys ser thr  
 1021/341  
 acg gcg gat tgc tgc gca tgc aag aca ccg acc cag gcg gcc agc ccc ctg gaa cgt cga  
 thr ala asp cys cys ala ser lys thr pro thr gln ala ala ser pro leu glu arg arg  
 1081/361  
 ttt acg tgc tgc tcc ccg gcc gtc gga tgc cga ttc cgc agc ttc ccg gtg cga cgg ctg  
 phe thr cys cys ser pro ala val gly cys arg phe arg ser phe pro val arg arg leu  
 1141/381  
 gcg ctg gga gca cgg aca tgc aga act ctc ggg gtt cgg cga acg tta tct cag tgg aat  
 ala leu gly ala arg thr ser arg thr leu gly val arg arg thr leu ser gln trp asn  
 1201/401  
 ctc agt cca cgc gcg caa cct agt tgt gca gtt act gtt gaa agc cac acc cat gcc agt  
 leu ser pro arg ala gln pro ser cys ala val thr val glu ser his thr his ala ser  
 1261/421  
 cca cgt atg gcc aag ttg gcc cga gta gtg ggc cta gta cag gaa gag caa cct agc gac  
 pro arg met ala lys leu ala arg val val gly leu val gln glu glu gln pro ser asp  
 1321/441  
 atg acg aat cac cca cgg tat tgc cca ccg ccg cag cag ccg gga acc cca ggt tat gct  
 met thr asn his pro arg tyr ser pro pro pro gln gln pro gly thr pro gly tyr ala  
 1381/461  
 cag ggg cag cag caa acg tac agc cag cag ttc gac tgg cgt tac cca ccg tcc ccg ccc  
 gln gly gln gln thr tyr ser gln gln phe asp trp arg tyr pro pro ser pro pro  
 1441/481  
 ccg cag cca acc cag tac cgt caa ccc tac gag gcg ttg ggt ggt acc cgg ccg ggt ctg  
 pro gln pro thr gln tyr arg gln pro tyr glu ala leu gly gly thr arg pro gly leu  
 1501/501  
 ata cct gcc gtg att ccg acc atg acc ccc cct cct gcc atg gtt cga caa cgc cct cgt  
 ile pro gly val ile pro thr met thr pro pro pro gly met val arg gln arg pro arg  
 1561/521  
 cca gcc atg ttg gcc atc gcc gcg gtg acc ata cct atg gtg tac gcc gcc atc gcc gcc  
 ala gly met leu ala ile gly ala val thr ile ala val val ser ala gly ile gly gly  
 1621/541  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 1681/561  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 1741/581  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 1801/601  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 1861/621  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 1921/641  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 1981/661  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 2041/681  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 2101/701  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 2161/721  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 2221/741  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 2281/761  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 2341/781  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala  
 2401/801  
 ggt acc acc cct cct cct ggt tta acc cca cca cca cca cca cca cca cca cca cca cca  
 ala

1/1	cat	cta	cac	cgc	tca	aca	gcc	ggg	cca	gac	31/11	gct	gcc	ggt	cgg	tgc	tgc	cga	gaa	ggc	ggt
his	leu	his	arg	ser	thr	ala	gly	pro	asp	ala	ala	ala	ala	gly	arg	cys	cys	arg	glu	gly	gly
61/21	gat	ccg	tgg	cga	gtt	gtt	cat	gtc	gcg	gcg	91/31	cac	cac	cgc	cga	cca	acg	ggt	gct	tgc	cat
asp	pro	trp	arg	val	val	his	val	ala	ala	ala	151/51	his	his	arg	arg	pro	thr	gly	ala	cys	his
121/41	ccg	tct	gac	caa	cgg	tag	ttc	gct	gct	gat	211/71	ctc	caa	aag	tct	caa	gcc	cac	cga	agc	agt
pro	ser	asp	gln	arg	AMB	phe	ala	ala	asp	leu	gln	lys	ser	gln	ala	his	arg	ser	ser		
181/61	cat	gaa	caa	qct	gcg	ttg	ggt	gct	att	gat	211/71	cgt	ggg	tgg	gat	cgg	ggt	ggc	ggt	cgc	cgc
his	glu	gln	ala	ala	leu	gly	ala	ile	asp	arg	gly	trp	asp	arg	gly	gly	gly	arg	arg		
241/81	ggt	ggc	cgg	ggg	gat	ggt	caa	cgg	ggc	cgg	271/91	gct	gag	gcc	ggt	ggg	cgg	cct	cac	cga	agc
gly	gly	arg	gly	asp	gly	his	pro	gly	arg	ala	glu	ala	gly	gly	pro	pro	his	arg	ser		
301/101	ggc	cga	gcg	ggt	ggc	gcg	aac	cga	cga	cct	331/111	gcg	gcc	cat	ccc	cgt	ctt	cgg	cag	cga	cga
gly	arg	ala	gly	gly	ala	asn	arg	arg	pro	ala	ala	his	pro	arg	leu	arg	gln	arg	arg		
361/121	att	ggc	cag	gct	gac	aga	ggc	att	caa	ttt	391/131	aat	gct	gcg	ggc	gct	ggc	cga	gtc	acg	gga
ile	gly	gln	ala	asp	arg	gly	ile	gln	phe	asn	ala	ala	gly	ala	gly	arg	val	thr	gly		
421/141	acg	gca	ggc	aag	gct	ggt	tac	cga	cgc	cgg	451/151	aca	tga	att	gcg	tac	ccc	gct	aac	gtc	gct
thr	ala	gly	lys	ala	gly	tyr	arg	arg	arg	thr	OPA	ile	ala	tyr	pro	ala	asn	val	ala		
481/161	gcg	cac	caa	tgt	cga	act	ctt	gat	ggc	ctc	511/171	gat	ggc	ccc	ggg	ggc	toc	gcg	gct	acc	caa
ala	his	gln	cys	arg	thr	leu	asp	gly	leu	asp	gly	pro	gly	gly	ser	ala	ala	thr	gln		
541/181	gca	gga	gat	ggt	cga	cct	ggg	tgc	cga	tgt	571/191	gct	ggc	tca	aat	cga	gga	att	gtc	cac	act
ala	gly	asp	gly	arg	pro	ala	cys	arg	cys	ala	gly	ser	asn	arg	gly	ile	val	his	thr		
601/201	ggt	agg	cga	ttt	ggt	gga	cct	gtc	cgg	agg	631/211	cga	cgt	cgg	aga	agt	ggt	gca	cga	gcc	ggt
gly	arg	arg	phe	gly	gly	pro	val	pro	arg	arg	arg	arg	arg	ser	gly	ala	arg	ala	gly		
661/221	cga	cat	gga	tgt	cgt	cgt	cga	cgg	cag	cat	691/231	gga	ggt	ggt	gag	ggt	gag	gag	caa	cga	cat
arg	his	gly	OPA	arg	arg	arg	pro	ala	pro	gly	ala	gly	gln	ala	ala	ala	gln	arg	tyr		
721/241	cat	ttt	cat	cat	cat	cat	cat	cat	cat	cat	751/251	ggt	ttt	cat	cat	cat	cat	cat	cat	cat	cat
pro	phe	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg	arg
781/261	cat	ggt	ggt	ggt	ggt	ggt	ggt	ggt	ggt	ggt	811/271	ggt	ggt	ggt	ggt	ggt	ggt	ggt	ggt	ggt	ggt
asp	gly	ala	ala	ala	ala	ala</															



169/185

1081/361  
tta cgt gct gct ccc cgg ccg tgg gat gcc gat tcc gca gct tcc cgg tgc gac gcc tgg  
leu arg ala ala pro arg pro ser asp ala asp ser ala ala ser arg cys asp gly trp  
1141/381  
cgc tgc gag cac gga cat cga gaa ctc tgc ggg ttc ggc gaa cgt tct ctc agt gga atc  
arg ser glu his gly his arg glu leu ser gly phe gly glu arg tyr leu ser gly ile  
1201/401  
tca gtc cac gcg cgc aac cta gtt ttg cag tta ctg ttg aaa gcc aca ccc atg cca gtc  
ser val his ala arg asn leu val val gln leu leu leu lys ala thr pro met pro val  
1261/421  
cac gca tgg cca agt tgg ccc gag tag tgg gcc tag tac agg aag agc aac cta gcg aca  
his ala trp pro ser trp pro glu AMB tip ala AMB tyr arg lys ser asn leu ala thr  
1321/441  
tga cga atc acc cac ggt att cgc cac cgc cgc agt agt agt agt gaa ccc cag gtt atg ctc  
OPA arg ile thr his gly ile arg his arg arg ser ser arg glu pro gln val met leu  
1381/461  
agg ggc agc agc aaa cgt aca gcc agc agt tgg agt ggc gtt acc cac cgt ccc cgc ccc  
arg gly ser ser lys arg thr ala ser ser ser thr gly val thr his arg pro arg pro  
1441/481  
cgc agc cca ccc agt acc gtc aac cct acg agg cgt tgg gtg gta ccc gcc cgg gtc tga  
arg ser gln pro ser thr val asn pro thr arg arg trp val val pro gly arg val OPA  
1501/501  
tac ctg gcg tga ttc cga cca tga cgc ccc ctc ctg gga tgg ttc gcc aac gcc ctc gtg  
tyr leu ala OPA phe arg pro OPA arg pro leu leu gly tip phe ala asn ala leu val  
1561/521  
cag gca tgt tgg cca tgc gcg cgg tga cga tag cgt tgg tgt tgg cgc gcc tgc gcg gcg  
gln ala cys tip pro ser ala arg OPA arg AMB arg tip tgg pro pro ala ser ala ala  
1621/541  
cgg ccg cat ccc tgg tcc ggt tca acc ggt cgt tca tca tca tca tca ggt cgc tgc acc agc  
arg pro his pro tip ser gly ser thr gly his pro pro ala pro ala ala ala gln trp  
1681/561  
ctc cca gag cgg cgc cca gcc tcc cgg tca tca tca tca tca tca ggt cgc tgc acc agc  
leu pro ala arg arg gln ala ser pro ala ala tca tca tca tca tca tca tca tca tca tca tca  
1741/581  
tgg cgg cca acc tgg tca cca gta tca tca tca tca tca tca tca tca tca tca tca tca tca  
trp arg pro arg trp tga pro val ser ser tca tca tca tca tca tca tca tca tca tca tca  
1801/601  
tca agt ggt tca ggt tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca  
tca arg ala pro ala tca pro tca tca tca tca tca tca tca tca tca tca tca tca tca tca  
1861/621  
tca cgg cgt acc cca acc cca acc tga tca tca tca tca tca tca tca tca tca tca tca tca  
ser arg tca tca pro tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca tca

170/185

1/1 31/11  
 atc tac acc gct caa cag ccg gcc cag acg ctg ccg gtc ggt gct gcc gag aag gcg gtg  
 ile tyr thr ala gln gln pro gly gln thr leu pro val gly ala ala glu lys ala val  
 61/21 91/31  
 atc cgt gcc gag ttg ttc atg tgc cgg cgc acc acc gcc gac caa cgg gtg ctt gcc atc  
 ile arg gly glu leu phe met ser arg arg thr thr ala asp gln arg val leu ala ile  
 121/41 151/51  
 cgt ctg acc aac ggt agt tgc ctg ctg atc tcc aac agt ctc aag ccc acc gaa gca gtc  
 arg leu thr asn gly ser ser leu leu ile ser lys ser leu lys pro thr glu ala val  
 181/61 211/71  
 atg aac aag ctg cgt tgg gtg cta ttg atc gtg ggt ggg atc ggg gtg gcg gtc gcc gcg  
 met asn lys leu arg trp val leu leu ile val giv gly ile gly val ala val ala ala  
 241/81 271/91  
 gtg gcc ggg ggg atg gtc acc cgg gcc ggg ctg agg cgg gtg gcc cgc ctc acc gaa gcg  
 val ala gly gly met val thr arg ala gly leu arg pro val gly arg leu thr glu ala  
 301/101 331/111  
 gcc gag cgg gtg gcg cga acc gac gac ctg cgg ccc atc ccc gtc ttc gcc agc gac gaa  
 ala glu arg val ala arg thr asp asp leu arg pro ile pro val phe gly ser asp glu  
 361/121 391/131  
 ttg gcc agg ctg aca gag gca ttc aat tta atg ctg cgg gcg ctg gcc gag tca cgg gaa  
 leu ala arg leu thr glu ala phe asn leu met leu arg ala leu ala glu ser arg glu  
 421/141 451/151  
 cgg cag gca agg ctg gtt acc gac gcc gga cat gaa ttg cgt acc cgg cta acg tgc ctg  
 arg gln ala arg leu val thr asp ala gly his glu leu arg thr pro leu thr ser leu  
 481/161 511/171  
 cgc acc aat gtc qaa ctc ttg atg gcc tgc atg gcc cgg ggg gct ccg cgg cta ccc aag  
 arg thr asn val glu leu leu met ala ser met ala pro gly ala pro arg leu pro lys  
 541/181 571/191  
 cag gag atg gtc gac cgc cgt gcc gat gtg ctg gct aac atc gag qaa ttg tcc aca ctg  
 gln glu met val asp leu arg ala asp val leu ala gln ile glu glu leu ser thr leu  
 601/201 631/211  
 gla ggt gat ttg gtg gac ctg tcc cga gcc cgc acc gga gag ggt gtc agt cgg cgg cgc aac gat atc  
 val gly asp leu val asp leu ser arg gly asp ala gly glu val val his glu pro val  
 661/221 691/231  
 gcc atg ggt gac gtc gtc gac cgc acc gga gag ggt gtc agt cgg cgg cgc aac gat atc  
 asp met ala asp val val asp arg ser leu ile ala val arg arg arg arg asn asp ile  
 721/241 751/251  
 ctt ttc gac gtc gag gtg att gac ttc gat att ttc acc ttc acc ggt ggt ttc ttc ggt  
 leu phe arg val glu val ile gly trp glu val ttc giv asp thr ala gly leu ser arg  
 781/261 811/271  
 atc ggt gtc acc gtc atc gtc acc ttc gtc acc ttc gtc acc ttc gtc acc ttc gtc acc  
 met ala leu asn leu met asp asn ala ala lys trp ser pro pro gly gly his val giv  
 841/281 871/291  
 gtc agt cgt acc gag gtc gat gtc gtc acc ttc gtc acc ttc gtc acc ttc gtc acc  
 val arg leu ser gln leu asp ala ser his ala ala leu val val ser asp asp gly pr  
 901/301 931/311  
 atc ggt gtc acc gtc atc gtc acc ttc gtc acc ttc gtc acc ttc gtc acc ttc gtc acc  
 ty thr pr val ala val arg arg arg arg arg arg arg arg arg arg arg arg arg arg

171/185

1021/341  
ggc gga ttg ctg cgc atc gaa gac acc gac cca ggc ggc cag ccc cct gga acg tcg att  
gly gly leu leu arg ile glu asp thr asp pro gly gly gln pro pro gly thr ser ile  
1081/361  
tac gtg ctg ctc ccc ggc cgt cgg atg ccg att ccg cag ctt ccc ggt gcg acg gct ggc  
tyr val leu leu pro gly arg arg met pro ile pro gln leu pro gly ala thr ala gly  
1141/381  
gct cgg agc acg gac atc gag aac tct cgg ggt tcc gcg aac gtt atc tca gtg gaa tct  
ala arg ser thr asp ile glu asn ser arg gly ser ala asn val ile ser val glu ser  
1201/401  
cag tcc acg cgc gca acc tag ttg tgc agt tac tgt tga aag cca cac cca tgc cag tcc  
gln ser thr arg ala thr AMB leu cys ser tyr cys OPA lys pro his pro cys gln ser  
1261/421  
acg cat ggc caa gtt ggc ccg aqt agt ggg cct agt aca gga aga gca acc tag cga cat  
thr his gly gln val gly pro ser ser gly pro ser thr gly arg ala thr AMB arg his  
1321/441  
gac gaa tca ccc acg gta ttc gcc acc gcc gca gca gcc ggg aac ccc agg tta tgc tca  
asp glu ser pro thr val phe ala thr ala ala ala ala gly asn pro arg leu cys ser  
1381/461  
ggg gca gca gca aac gta cag cca gca gtt cga ctg gcg tta ccc acc gtc ccc gcc ccc  
gly ala ala ala asn val gln pro ala val arg leu ala leu pro thr val pro ala pro  
1441/481  
gca gcc aac cca gta ccg tca acc cta cga ggc gtt ggg tgg tac ccg gcc ggg tct gat  
ala ala asn pro val pro ser thr leu arg gly val gly trp tyr pro ala gly ser asp  
1501/501  
acc tgg cgt gat tcc gac cat gac gcc ccc tcc tgg gat ggt tgg cca acg ccc tgg tgc  
thr trp arg asp ser asp his asp ala pro ser trp asp gly ser pro thr pro ser cys  
1561/521  
agg cat gtt gcc cat cgg cgc ggt gac gat agc ggt ggt gtc cgc cgg cat cgg cgg cgt  
arg his val gly his arg arg gly asp asp ser gly gly val arg arg his arg arg arg  
1621/541  
ggc cgt atc cct ggt cgg gtt cca ccg ggc acc cgc cgg ccc cca cgt cgg ccc agt ggc  
gly arg ile pro gly arg val gln pro gly thr arg arg pro gln arg arg pro ser gly  
1681/561  
tac cca cct acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc acc  
cys gln arg arg ala lys his pro arg ser lys his ala ala gly val gly arg thr gly  
1741/581  
glt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt  
gly gly gln gly gly ala gln cys arg his val gly arg arg arg arg arg arg arg arg  
1801/601  
ggt  
gly gly leu arg his his ser val cys arg gly ala arg leu arg gln gln pro arg arg  
1861/621  
ggt  
arg  
1921/641  
ggt  
arg  
1981/661  
ggt  
arg  
2041/681  
ggt  
arg  
2101/701  
ggt  
arg  
2161/721  
ggt  
arg  
2221/741  
ggt  
arg  
2281/761  
ggt  
arg  
2341/781  
ggt  
arg  
2401/801  
ggt  
arg  
2461/821  
ggt  
arg  
2521/841  
ggt  
arg  
2581/861  
ggt  
arg  
2641/881  
ggt  
arg  
2701/901  
ggt  
arg  
2761/921  
ggt  
arg  
2821/941  
ggt  
arg  
2881/961  
ggt  
arg  
2941/981  
ggt  
arg  
3001/1001  
ggt  
arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg arg

171/185

172/185

Séquence codante Rv0983 prédite par Cole et al., 1998 (Nature 393:537-544) et contenant seq60A:

```

1/1                               31/11
atg gcc aag ttg gcc cga gta gtg ggc cta gta cag gaa gag caa cct agc gac atg acg
Met ala lys leu ala arg val val gly leu val gln glu glu gln pro ser asp met thr
61/21                               91/31
aat cac cca cgg tat tgg cca ccg ccg cag cag ccg gga acc cca ggt tat gat cag ggg
asn his pro arg tyr ser pro pro pro gln gln pro gly thr pro gly tyr ala gln gly
121/41                               151/51
cag cag caa acg tac agc cag cag ttc gac tgg cgt tat cca ccg tcc ccg ccc ccg cag
gln gln gln thr tyr ser gln gln phe asp trp arg tyr pro pro ser pro pro pro gln
181/61                               211/71
cca acc cag tac cgt caa ccc tac gag gcg ttg ggt ggt acc ccg ccg ggt ctg ata cct
pro thr gln tyr arg gln pro tyr glu ala leu gly gly thr arg pro gly leu ile pro
241/81                               271/91
ggc gtg att ccg acc atg acg ccc cct cct ggg atg gtt cgc caa cgc cct cgt gca ggc
gly val ile pro thr met thr pro pro pro gly met val arg gln arg pro arg ala gly
301/101                               331/111
atg ttg gcc atc ggc ggc gtg acg ata gcg gtg gtg tcc gcc ggc atc ggc ggc ggc gcc
met leu ala ile gly ala val thr ile ala val val ser ala gly ile gly gly ala ala
361/121                               391/131
gca tcc ctg gtc ggg ttc aac cgg gca ccc gcc ggc ccc acc ggc ggc cca gtg gct gcc
ala ser leu val gly phe asn arg ala pro ala gly pro ser gly gly pro val ala ala
421/141                               451/151
agc gcg gcg cca agc atc ccc gca gca aac atg ccg ccg ggg tgg gtc gaa cag gtg gcg
ser ala ala pro ser ile pro ala ala asn met pro pro gly ser val glu gln val ala
481/161                               511/171
gcc aag gtg gtg ccc agt gtc gtc atg ttg gaa acc gat ctg ggc cgc cag tgg gag gag
ala lys val val pro ser val val met leu glu thr asp leu gly arg gln ser glu glu
541/181                               571/191
ggc tcc gcc atc att ctg tct gcc gag gga ctg atc ttg acc aac aac cag gtg atc ccg
gly ser gly ile ile leu ser ala glu gly leu ile leu thr asn asn his val ile ala
601/201                               631/211
ggg gcc gcc aag cct cca ctg ggc agt cca cca cca aac acc acc ata acc tta tta gca
ala ala ala lys pro pro leu gly ser pro pro pro gly thr thr val thr phe ser asp
661/221                               691/231
ggg cgg acc cca cct tta acc gtg atg gaa gat gat acc acc acc gta acc gta gta gta
gly arg thr ala pro pro thr val val gly ala asp pro thr acc acc ala val val
721/241                               751/251
gtt gtt acc gta atc acc gta acc gta acc gta acc gta acc gta acc gta acc gta acc
arg val gln gly val ser gly leu thr pro ile ser leu gly ser ser asp leu arg
781/261                               811/271
gtt gtt acc cca gta acc gaa atc tta acc acc gta acc gta acc gta acc gta acc gta
val gly gln pro val leu ala ile gly arg pro leu gly leu gly gly thr val thr thr

```

173/185

841/281	871/291
ggg atc gtc agc gct ctc aac cgt cca gtg	tcg acg acc ggc gag gcc ggc aac cag aac
gly ile val ser ala leu asn arg pro val	ser thr thr gly glu ala gly asn gln asn
901/301	931/311
acc gtg ctg gac gcc att cag acc gac gcc	gcg atc aac ccc ggt aac tcc ggg ggc gcg
thr val leu asp ala ile gln thr asp ala	ala ile asn pro gly asn ser gly gly ala
961/321	991/331
ctg gtg aac atg aac gct caa ctc gtc gga	gtc aac tcg gcc att gcc acg ctg gcc gcg
leu val asn met asn ala gln leu val gly	val asn ser ala ile ala thr leu gly ala
1021/341	1051/351
gac tca gcc gat gcg cag agc gcc tcg atc	ggt ctc ggt ttt gcg att cca gtc gac cag
asp ser ala asp ala gln ser gly ser ile	gly leu gly phe ala ile pro val asp gln
1081/361	1111/371
gcc aag cgc atc gcc gac gag ttg atc agc	acc gcc aag gcg tca cat gcc tcc ctg ggt
ala lys arg ile ala asp glu leu ile ser	thr gly lys ala ser his ala ser leu gly
1141/381	1171/391
gtg cag gtg acc aat gac aaa gac acc ctg	gac gcc aag atc gtc gaa gta gtg gcc ggt
val gln val thr asn asp lys asp thr leu	gly ala lys ile val glu val val ala gly
1201/401	1231/411
ggt gct gcc gcg aac gct gga gtg ccg aag	ggg gtc att gtc acc aag gtc gac gac cgc
gly ala ala ala asn ala gly val pro lys	gly val val val thr lys val asp asp arg
1261/421	1291/431
ccg atc aac agc gcg gac gcg ttg gtt gcc	gcc gtg ccg tcc aaa gcg ccg gcc gcc acg
pro ile asn ser ala asp ala leu val ala	ala val arg ser lys ala pro gly ala thr
1321/441	1351/451
gtg gcg cta acc ttt cag gat ccc tcg gcc	ggt agc cgc acc gtg caa gtc acc ctc gcc
val ala leu thr phe gln asp pro ser gly	gly ser arg thr val gln val thr leu gly
1381/461	
aag gcg gag cag tga	
lys ala glu gln opa	

SEQ ID N° 490 (suite 1)

FIGURE 490 (suite 1)

174/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv0983

1/1  
tga gcc agc tcg acg cgt cgc acg ctg agc tgg tgg ttt ccg acc gcg gcc cgg gca ttc  
OPA ala ser ser thr arg arg thr leu ser trp trp phe pro thr ala ala arg ala phe  
61/21  
ccg tgc agg agc gcc gtc tgg tgt ttg aac ggt ttt acc ggt cgg cat cgg cac ggg cgt  
pro cys arg ser ala val trp cys leu asn gly phe thr gly arg his arg his gly arg  
121/41  
tgc cgg gtt cgg gcc tcg ggt tgg cga tcg tca aac agg tgg tgc tca acc acg gcg gat  
cys arg val arg ala ser gly trp arg ser ser asn arg trp cys ser thr thr ala asp  
181/61  
tgc tgc gca tcg aag aca ccg acc cag gcg gcc agc ccc ctg qaa cat cga ttt acg tgc  
cys cys ala ser lys thr pro thr gln ala ala ser pro leu glu arg arg phe thr cys  
241/81  
tgc tcc ccg gcc gtc gga tgc cga ttc cgc agc ttc ccg gtg cga cgg ctg gcg ctc gga  
cys ser pro ala val gly cys arg phe arg ser phe pro val arg arg leu ala leu gly  
301/101  
gca cgg aca tcg aga act ctc ggg gtt cgg cga acc tta tct cag tgg aat ctc agt cca  
ala arg thr ser arg thr leu gly val arg arg thr leu ser gln trp asn leu ser pro  
361/121  
cgc gcg caa cct agt tgt gca gtt act gtt gaa agc cac acc cat gcc agt cca cgc atg  
arg ala gln pro ser cys ala val thr val glu ser his thr his ala ser pro arg met  
421/141  
gcc aag ttg gcc cga gta gtg ggc cta gta cag gaa gag caa cct agc gac atg acg aat  
ala lys leu ala arg val val gly leu val gln glu glu gln pro ser asp met thr asn  
481/161  
cac cca cgg tat tcg cca ccg ccg cag cag ccg gaa acc cca ggt tat gat cag ggg cag  
his pro arg tyr ser pro pro pro gln gln pro gly thr pro gly tyr ala gln gly gln  
541/181  
cag caa acg tac agc cag cag ttc gac tgg cgt tat cca ccg tcc ccg ccc cag cca  
gln gln thr tyr ser gln gln phe asp trp arg tyr pro pro ser pro pro pro gln pro  
601/201  
acc cag tac cgt caa ccc tac gag gcg ttg ggt gat acc ccg ccg ggt ctg ata cct ggt  
thr gln tyr arg gln pro tyr glu ala leu gly gly thr arg pro gly leu ile pro gly  
661/221  
atg att cca acc atg acc ccc cct cct ggt atg gtt cca acc cgt cct ggt ggt ggt  
val ile pro thr met thr pro pro pro gly met val thr thr thr thr thr thr thr thr  
721/241  
ttg gca atg ggt cag gca acc cta gag gta gta ttc cca acc cca acc cca acc cca  
leu ala ile gly ala met thr ala ala val val met cca acc cca acc cca acc cca  
781/261  
tgc cgt cgt ggt ttc acc arg cca cca pro arg ttc cca acc cca acc cca acc cca  
ser leu val gly phe acc arg ala pro ala gly pro ser gly gly pro val ala ala cca  
841/281  
gtg ggt cca acc atg cca acc acc acc acc acc acc acc acc acc acc acc acc acc  
ala ala pro met ala met ala ala acc met thr pro acc acc val ala thr met ala thr

SEQUENCE 1

175/185

901/301  
 aag gtg gtg ccc agt gtc gtc atg ttg gaa acc gat ctg ggc cgc cag tcc gag gag ggc  
 lys val val pro ser val val met leu glu thr asp leu gly arg gln ser glu glu gly  
 961/321  
 tcc ggc atc att ctg tct gcc gag ggg ctg atc ttg acc aac aac cac gtg atc gcg gcg  
 ser gly ile ile leu ser ala glu gly leu ile leu thr asn asn his val ile ala ala  
 1021/341  
 gcc gcc aag cct ccc ctg ggc agt ccg ccg ccg aaa acg acg gta acc ttc tct gac ggg  
 ala ala lys pro pro leu gly ser pro pro pro lys thr thr val thr phe ser asp gly  
 1091/361  
 cgg acc gca ccc ttc acg gtg gtg ggg gct gac ccc acc agt gat atc gcc gtc gtc cgt  
 arg thr ala pro phe thr val val gly ala asp pro thr ser asp ile ala val val arg  
 1141/381  
 gtt caa ggc gtc tcc ggg ctc acc ccg atc tcc ctg ggt tcc tcc tcc gag ctg agg gtc  
 val gln gly val ser gly leu thr pro ile ser leu gly ser ser ser asp leu arg val  
 1201/401  
 ggt cag ccg gtg ctg gcg atc ggg tcc ccg ctc ggt ttg gag ggc acc gtg acc acg ggg  
 gly gln pro val leu ala ile gly ser pro leu gly leu glu gly thr val thr thr gly  
 1261/421  
 atc gtc agc gct ctc aac cgt cca gtg tcc acg acc ggc gag gcc ggc aac cag aac acc  
 ile val ser ala leu asn arg pro val ser thr thr gly glu ala gly asn gln asn thr  
 1321/441  
 gtg ctg gac gcc att cag acc gac gcc gcg atc aac ccc ggt aac tcc ggg ggc gcg ctg  
 val leu asp ala ile gln thr asp ala ala ile asn pro gly asn ser gly gly ala leu  
 1381/461  
 gtg aac atg aac gct caa ctc gtc gga gtc aac tcc gcc att gcc acg ctg ggc gcg gac  
 val asn met asn ala gln leu val gly val asn ser ala ile ala thr leu gly ala asp  
 1441/481  
 tca gcc gat gcg cag agc ggc tcc atc ggt ctc ggt ttt gcg att cca gtc gac cag gcc  
 ser ala asp ala gln ser gly ser ile gly leu gly phe ala ile pro val asp gln ala  
 1501/501  
 aag cgc atc gcc gac gag ttg atc agc acc gcc aag gcg tca cat gcc tcc ctg ggt gcg  
 lys arg ile ala asp glu leu ile ser thr gly lys ala ser his ala ser leu gly val  
 1561/521  
 cag gtg acc aat gac aac gac acc ctg gcg ctc ggt ttt gcg att cca gtc gac cag gcc  
 gln val thr asn asp lys asp thr leu gly ala lys ala val ala val val ala ile gly  
 1621/541  
 ctc gcc ccg aac gct gca gag ccg aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ala ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 1681/561  
 atc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ile asn ser ala asp ala leu val ala ala val asn ser lys ala pro gly ala thr val  
 1741/581  
 ctc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ala ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 1801/601  
 atc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ile ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 1861/621  
 ctc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ala ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 1921/641  
 atc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ile ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 1981/661  
 ctc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ala ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 2041/681  
 atc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ile ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 2101/701  
 ctc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ala ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 2161/721  
 atc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ile ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 2221/741  
 ctc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ala ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 2281/761  
 atc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ile ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 2341/781  
 ctc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ala ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro  
 2401/801  
 atc aac acc gcg gac ggc ttg att gtc aac gtc ggt ttt gcg att cca gtc gac cag gcc  
 ile ala ala asn ala gly val pro lys gly val val val thr thr lys val asp asp arg pro

Fragment amplifié par PCR d'après les similarités de séquence avec une sérine protéase de la famille HtrA de *E. coli* (création du site SnaBI à l'extrémité 3') et sous cloné dans le vecteur pJVEDa:

1/1	31/11																
gat ccg gcg ggg cgg gtg tgg gcg cag gcg	trp ctg gcg gtc acg gcg gtg cgg gcg gtg																
asp pro ala gly arg val ser ala gln ala	trp leu ala val thr ala val arg ala val																
61/21	91/31																
ccg ccg ggc tgt ggg gcg ccg gcg gcg gcg	gtg gca atg gcg gga acg gcg ccg atg cca																
pro pro gly cys gly ala pro ala ala ala	val ala met ala gly thr ala pro met pro																
121/41	151/51																
aca tgg tca gcg gtg gag acg gtg gcc tgg	gcg gtg ccg gtg gcg gtg gcg gat gcc tct																
thr ser ser ala val glu thr val ala ser	ala val pro val ala val ala asp gly ser																
181/61	211/71																
acg gcg acg gcg ggg ccg gcg gac acg gcg	gac aag gcg caa tgg gcc tgg gcg gcg gcg																
thr ala thr ala gly pro ala asp thr ala	asp lys ala gln ser ala ser ala ala ala																
241/81	271/91																
ccg gcg gcg acg ggg gcc agg gcg gcg ccg	gcc gcg gac tgt ggg gta ctg gcg gcg ccg																
pro ala ala thr gly ala arg ala ala pro	ala ala asp cys gly val leu ala ala pro																
301/101	331/111																
gcg gac acg gcg ggc aag gcg gtg gta ccg	ggg gcc cac cgg tgc ccg gtc agg cag gca																
ala asp thr ala gly lys ala val val pro	gly ala his arg cys pro val arg gln ala																
361/121	391/131																
tgg gcg ccg ccg gtg gcg ccg gtg gcc tga	tgg gca acg gcg ggg ccg gcg gcg acg gcg																
trp ala pro arg val ala pro val gly opa	ser ala thr ala gly pro ala ala thr ala																
421/141	451/151																
gtg tgg gcg cgt ccg gcg ggg tgg ccg gag	tag gcg gtg ccg gcg gga acg cca tgc tga																
val ser ala arg pro ala gly ser pro glu	amb ala val pro ala gly thr pro cys opa																
481/161	511/171																
tgg ggc acg gcg gcg ccg gcg gcg ccg gcg	gag aca gca gtt tgg cta atg gcg cgg ccg																
ser gly thr ala ala pro ala ala pro ala	glu thr ala val ser leu met ala arg pro																
541/181	571/191																
gcg gcg ccg gcg gtg ccg gag gcc acc tct	tgg gca atg gcg ggt ccg gcg gcc acg gcg																
ala ala arg ala val pro glu gly thr ser	ser ala met ala gly pro ala ala thr ala																
601/201	631/211																
gag ccg tca tgg ccg gca aca ccg gta tgg	gtg gcg ccg gcg gcg tgg gtg ggg acg cca																
glu pro ser arg pro ala thr pro val ser	val ala pro ala ala ser val gly thr pro																
661/221	691/231																
gcc tga tgg gcc agg gtg gcg ccg gcg gtg	ccg gcg ggg acc gcg ccg gag cct tgg ttg																
gly opa ser ala thr val ala pro ala val	pro ala gly thr ala pro glu pro trp leu																
721/241	751/251																
gcc gtg acc tgg gcc ccg gtg gga acg ggg	gcg ctg acc gga agc tat acc gca acc gcg																
ala val thr ala gly pro val gly thr gly	ala leu ala ala ser thr thr ala thr ala																
781/261	811/271																
gag acc ggg gcc ccg gca ccg ccg gaa cca	tgc acc ccg cca tta ccg tat tga ccg																
ala thr ala pro pro ala pro ala glu his	cys ala arg ala ala ala asp tip opa acc																
841/281	871/291																
ala tgg tta tgg cca ccg gcc acc tgg gcg	aca acc gcc acc ccg tat acc gcc tat cca																
leu cys ser val his pro ala ala pro ala	thr pro ala ala pro ala ala ser asp gln																
901/301	931/311																
gga ggg tta ggt tgg ggt tgg ccg gta tgg	cca tgg ttt gta ttt ggt ggt tgg acc acc																
agg gly phe acc tga arg ser gly ala ttt	pro ser ala ala leu ala ala tgg thr thr																
961/321	991/331																
tta tta tta tta tta tta tta tta tta tta	tta tta tta tta tta tta tta tta tta tta																
val val amb leu ala pro ala ala tta tta	tta tta tta tta tta tta tta tta tta tta																
991/341	1000/351																
tta tta tta tta tta tta tta tta tta tta	tta tta tta tta tta tta tta tta tta tta																
tta tta tta tta tta tta tta tta tta tta	tta tta tta tta tta tta tta tta tta tta																



177/185

1/1  
 atc cgg cgg ggc ggg tgt cgg cgc aag cgt ggc tgg cgg tca cgg cgg tgc ggg cgg tgc  
 ile arg arg gly gly cys arg arg arg arg gly trp arg ser arg arg cys gly arg cys  
 61/21  
 cgc cgg gct gtg ggg cgc cgg cgg cgg cgg tgg caa tgg cgg gaa cgg cgc cga tgc caa  
 arg arg ala val gly arg arg arg arg arg trp gln trp arg glu arg arg arg cys gln  
 121/41  
 cat cgt cag cgg tgg aga cgg tgg cct cgg cgg tgc cgg tgg cgg tgg cgg atg gct cta  
 his arg gln arg trp arg arg trp pro arg arg cys arg trp arg trp arg met ala leu  
 181/61  
 cgg cga cgg cgg ggc cgg cgg aca cgg cgg aca agg cgc aat cgg cct cgg cgg cgg cgc  
 arg arg arg arg gly arg arg thr arg arg thr arg arg asn arg pro arg arg arg arg  
 241/81  
 cgg cgg cga cgg ggg cca ggg cgg cgc cgg cgg cgg act gtg aag tac tgg cgg cgc cgg  
 arg arg arg arg gly pro gly arg arg arg pro arg thr val gly tyr trp arg arg arg  
 301/101  
 cgg aca cgg cgg gca agg cgg tgg tac cgg ggg ccc acc gct gcc cgg tca ggc agg cat  
 arg thr arg arg ala arg arg trp tyr arg gly pro thr ala ala arg ser gly arg his  
 361/121  
 ggg cgc cgc ggg tgg cgt cgg tgg gct gat cgg caa cgg cgg cgg cgg cga cgg cgg  
 gly arg arg gly trp arg arg trp ala asp arg gln arg arg gly arg arg arg arg arg  
 421/141  
 tgt cgg cgc gtc cgg cgg ggt cgc cgg agt agg cgg tgc cgg cgg gaa cgc cat gct gat  
 cys arg arg val arg arg gly arg arg ser arg arg cys arg arg glu arg his ala asp  
 481/161  
 cgg gca cgg cgg cgc cgg cgg cgc cgg cgg aga cag cag ttt cgc taa tgg cgc gcc cgg  
 arg ala arg arg arg arg arg arg arg arg gln gln phe arg och trp arg gly arg  
 541/181  
 cgg cgc ggg cgg tgc cgg agg gca cct ctt cgg caa tgg cgg gtc cgg cgg cca cgg cgg  
 arg arg gly arg cys arg arg ala pro leu arg gln trp arg val arg arg pro arg arg  
 601/201  
 agc cgt cac gcc cgg caa cac cgg tat cgg tgg cgg cgg cgt cgt cgg tgg gga cgc cag  
 ser arg his gly arg gln his arg tyr arg trp arg arg arg arg arg trp gly arg gln  
 661/221  
 gct gat cgg cca cgg tgg cgt cgg cgg tgc cgg cgg gga cgg cgg cgg ago ctt cgt tgg  
 ala asp arg pro arg trp arg arg arg cys arg arg gly pro arg arg ser leu gly trp  
 721/241  
 cgg tga cgg cgg gcc cgg tgg gaa cgg cgg cgc tga cgg cca gct ata cgg caa cgg cgg  
 pro opa arg arg ala arg trp glu arg gly arg trp arg pro ala ile arg gln arg arg  
 781/261  
 cga cgg cgg cgg cgg cgg cgg cgg cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt  
 arg arg arg pro arg his arg arg arg arg arg arg arg arg arg arg arg arg arg arg  
 841/281  
 ttt gtt cgt ttc acc cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt  
 phe val arg cys thr arg pro thr arg arg his arg pro thr arg leu ala pro ile arg  
 901/301  
 tag ggt ttc gct gcc ggt cca ggt cat ggt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt  
 glu gly phe gly ala arg pro gly his gly his pro leu ser trp arg ser gly leu arg  
 961/321  
 tgg tgt aga ala atc cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt cgt  
 trp cys arg lys ala leu pro pro gly thr leu arg leu gly gln thr leu ala ala thr  
 1021/341  
 cgt  
 arg  
 1081/401

178/185

1/1 31/11  
 tcc gcc ggg gcg ggt gtc gcc gca gcc gtg gct gcc ggt cac gcc ggt gcg gcc ggt gcc  
 ser gly gly ala gly val gly ala gly val ala gly gly his gly gly ala gly gly ala  
 61/21 91/31  
 gcc ggg ctg tgg gcc gcc gcc gcc gcc ggt gcc aat gcc ggg aac gcc gcc gat gcc aac  
 ala gly leu trp gly ala gly gly gly gly gly asn gly gly asn gly ala asp ala asn  
 121/41 151/51  
 atc gtc agc ggt gga gac ggt gcc ctc gcc ggt gcc ggt gcc gga tgg ctc tac  
 ile val ser gly gly asp gly gly leu gly gly ala gly gly gly gly gly trp leu tyr  
 161/61 211/71  
 gcc gac gcc ggg gcc gcc gga cac gcc gga caa gcc gca atc gcc ctc gcc gcc gcc gcc  
 gly asp gly gly ala gly gly his gly gly gln gly ala ile gly leu gly gly gly ala  
 241/81 271/91  
 gcc gcc gac ggg gcc cag gcc gcc gcc gcc cgc gga ctg tgg ggt act gcc gcc gcc gcc  
 gly gly asp gly gly gln gly gly ala gly arg gly leu trp gly thr gly gly ala gly  
 301/101 331/111  
 gga cac gcc ggg caa gcc ggt ggt acc gcc gcc gcc ccc ctg ccc ggt caq gca gcc atg  
 gly his gly gly gln gly gly gly thr gly gly pro pro leu pro gly gln ala gly met  
 361/121 391/131  
 gcc gcc gcg ggt gcc gcc ggt ggg ctg atc gcc aac gcc ggg gcc gcc gcc gcc gcc  
 gly ala ala gly gly ala gly gly leu ile gly asn gly gly ala gly gly asp gly gly  
 421/141 451/151  
 gtc gcc gcg tcc gcc ggg gtc gcc gga gta gcc ggt gcc gcc ggg aac gcc atg ctg atc  
 val gly ala ser gly gly val ala gly val gly gly ala gly gly asn ala met leu ile  
 481/161 511/171  
 ggg cac gcc gcc gcc gcc gcc gcc gcc gga gac agc agt ttc gct aat gcc gcg gcc gcc  
 gly his gly gly ala gly gly ala gly gly asp ser ser phe ala asn gly ala ala gly  
 541/181 571/191  
 gcc gcg gcc ggt gcc gga ggg cac ctc ttc gcc aat gcc ggg tcc gcc gcc cac gcc gga  
 gly ala gly gly ala gly gly his leu phe gly asn gly gly ser gly gly his gly gly  
 601/201 631/211  
 gcc gtc acg gcc gcc aac acc ggt atc ggt gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 ala val thr ala gly asn thr gly ile gly gly ala gly gly val gly gly asp ala arg  
 661/221 691/231  
 ctg atc gcc cac ggt gcc gcc gcc ggt gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 leu ile gly his gly gly ala gly gly ala gly gly asp arg ala gly ala leu val gly  
 721/241 751/251  
 cgt gac gcc ggg gcc ggt ggg aac ggg gcc gct gcc gcc cag ata tac gcc aac gcc gcc  
 arg asp gly gly pro gly gly asn gly gly ala gly gly gly gln leu tyr gly asn gly gly  
 781/261 811/271  
 gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 asp gly ala pro gly thr gly gly thr leu ala ala val ala gly leu val thr ala  
 841/281 871/291  
 ttg ttc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 leu phe gly ala pro gly gln pro gly asp thr gly gln pro gly ala ala ala ala ala  
 901/301 931/311  
 gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 arg val ser val pro val arg gly met ala ile ala cpa ala gly asp leu asp tyr val  
 961/321 991/331  
 ggt ata caa aaa t  
 gly val ala gly ser gly arg pro asp pro thr thr thr thr thr thr thr thr thr thr thr  
 1021/341 1051/351  
 gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 thr  
 1081/361 1111/371  
 gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 thr  
 1141/381 1171/391  
 gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc  
 thr  
 1201/401 1231/411

179/185

Séquence codante Rv0125 prédite par Cole et al., 1998 (Nature 393:537-544) et contenant seq50A:

```

1/1                               31/11
atg agc aat tgc cgc cgc cgc tca ctc agg tgg tca tgg ttg ctg agc gtg ctg gct gcc
Met ser asn ser arg arg arg ser leu arg trp ser trp leu leu ser val leu ala ala
61/21                               91/31
gtc ggg ctg gcc ctg gcc acg gcg ccg gcc cag gcg gcc ccg ccg gcc ttg tgc cag gac
val gly leu gly leu ala thr ala pro ala gln ala ala pro pro ala leu ser gln asp
121/41                               151/51
egg ttc gcc gac ttc ccc gcg ctg ccc ctc gac ccg tcc gcg atg gtc gcc caa gtg ggg
arg phe ala asp phe pro ala leu pro leu asp pro ser ala met val ala gln val gly
181/61                               211/71
cca cag gtg gtc aac atc aac acc aaa ctg gcc tac aac aac gcc gtg gcc gcc ggg acc
pro gln val val asn ile asn thr lys leu gly tyr asn asn ala val gly ala gly thr
241/81                               271/91
ggc atc gtc atc gat ccc aac ggt gtc gtg ctg acc aac aac cac gtg atc gcg gcc gcc
gly ile val ile asp pro asn gly val val leu thr asn asn his val ile ala gly ala
301/101                               331/111
acc gac atc aat gcg ttc agc gtc gcc tcc gcc caa acc tac gcc gtc gat gtg gtc ggg
thr asp ile asn ala phe ser val gly ser gly gln thr tyr gly val asp val val gly
361/121                               391/131
tat gac cgc acc cag gat gtc gcg gtg ctg cag ctg cgc ggt gcc ggt gcc ctg ccg tgc
tyr asp arg thr gln asp val ala val leu gln leu arg gly ala gly gly leu pro ser
421/141                               451/151
gcg gcg atc ggt gcc gcc gtc gcg gtt ggt gag ccc gtc gtc gcg atg gcc aac agc ggt
ala ala ile gly gly gly val ala val gly glu pro val val ala met gly asn ser gly
481/161                               511/171
ggg cag gcc gga acg ccc cgt gcg gtg cct gcc agc gtg gtc gcg ctc gcc caa acc gtg
gly gln gly gly thr pro arg ala val pro gly arg val val ala leu gly gln thr val
541/181                               571/191
cag gcg tgc gat tgc ctg acc ggt gcc gaa gag aca ttg aac ggg ttg atc cag ttc gat
gln ala ser asp ser leu thr gly ala glu glu thr leu asn gly leu ile gln phe asp
601/201                               631/211
gac gcg atc cag acc ggt gat tgc gcc ggg ccc gtc gtc aac gcc cta gga cag gtg gtc
ala ala ile gln pro gly asp ser gly gly pro val val val asn gly leu gly gln val val
661/221                               691/231
ggt atg aac acc gcc gcg tcc gat aac ttc cag ctg tcc cag ggt gcc cag gga ttc gcc
gly met asn thr ala ala ser asp asn phe gln leu ser gln gly gly gln gly phe ala
721/241                               751/251
att tgc atc ggc cag gga atc gcg atc agc gac cag atc ggt tgc ggt ggt ttc ccc
ile pro ile gly gln ala met ala ile ala gly gln ile arg ser gly tyr gly ser pro
781/261                               811/271
atg gtc gat atc ggc ggt acc gcc ttc gtc aac ttc gtc atc gtc gtc gtc gtc gtc gtc
thr val asp ile gly pro thr ala phe leu gly leu gly val val asp asp asn gly asn
841/281                               871/291
acc gga atc gtc atc ggc ggt gtc ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt
gly ala arg val gln arg val val gly ser ala pro ala ala ser leu gly ile ser thr
901/301                               931/311
atg atc gtc atc ggc ggt gtc ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt ggt
gly arg val ile thr ala val asp gly ala pro ile arg ser ala thr ile ser ala thr

```

180/185

ORF d'après Cole et al., 1998 (Nature 393:537-544) et contenant Rv0125;

[illegible]

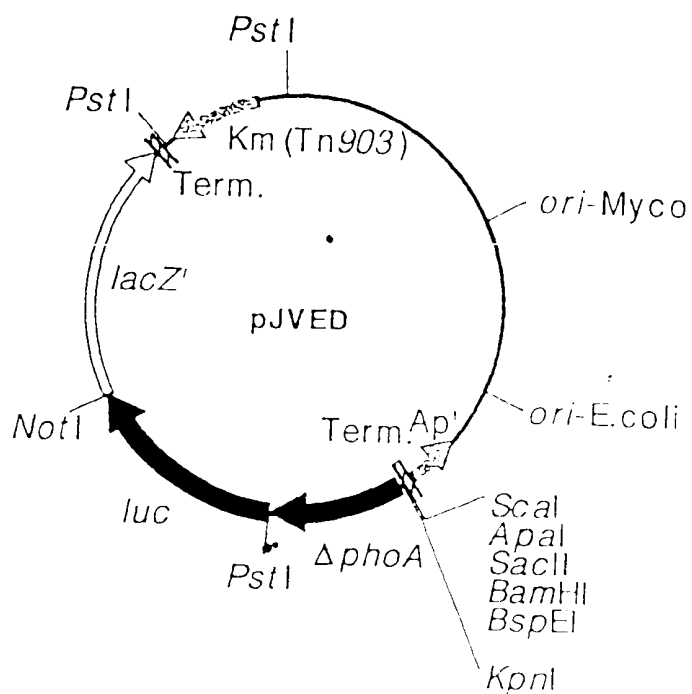
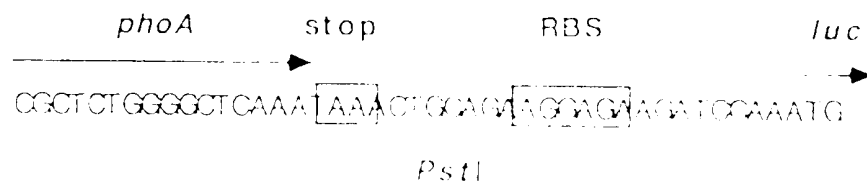
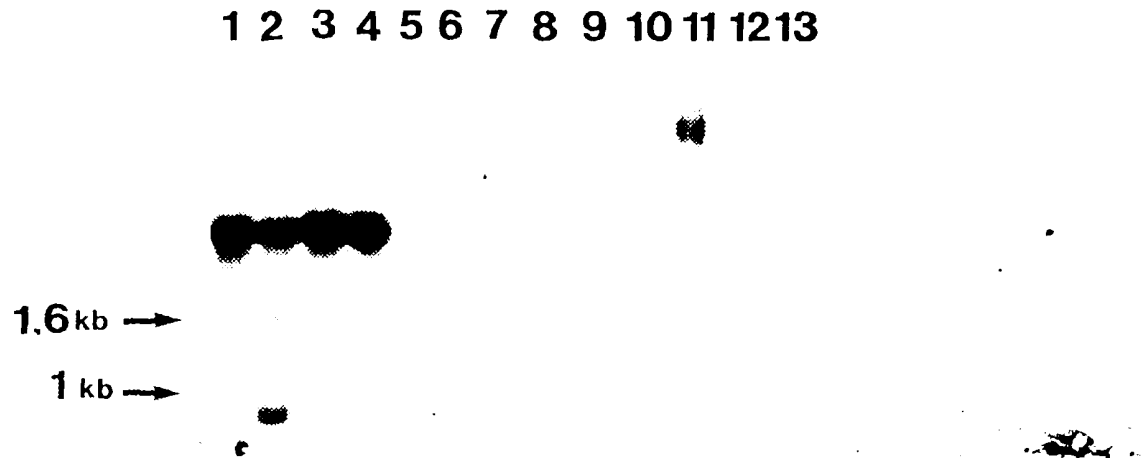


FIGURE 51A



182/185

Expériences d'hybridation moléculaire d'une sonde spécifique  
du DP428 sur l'ADN génomique de différentes espèces de  
mycobactéries



183/185

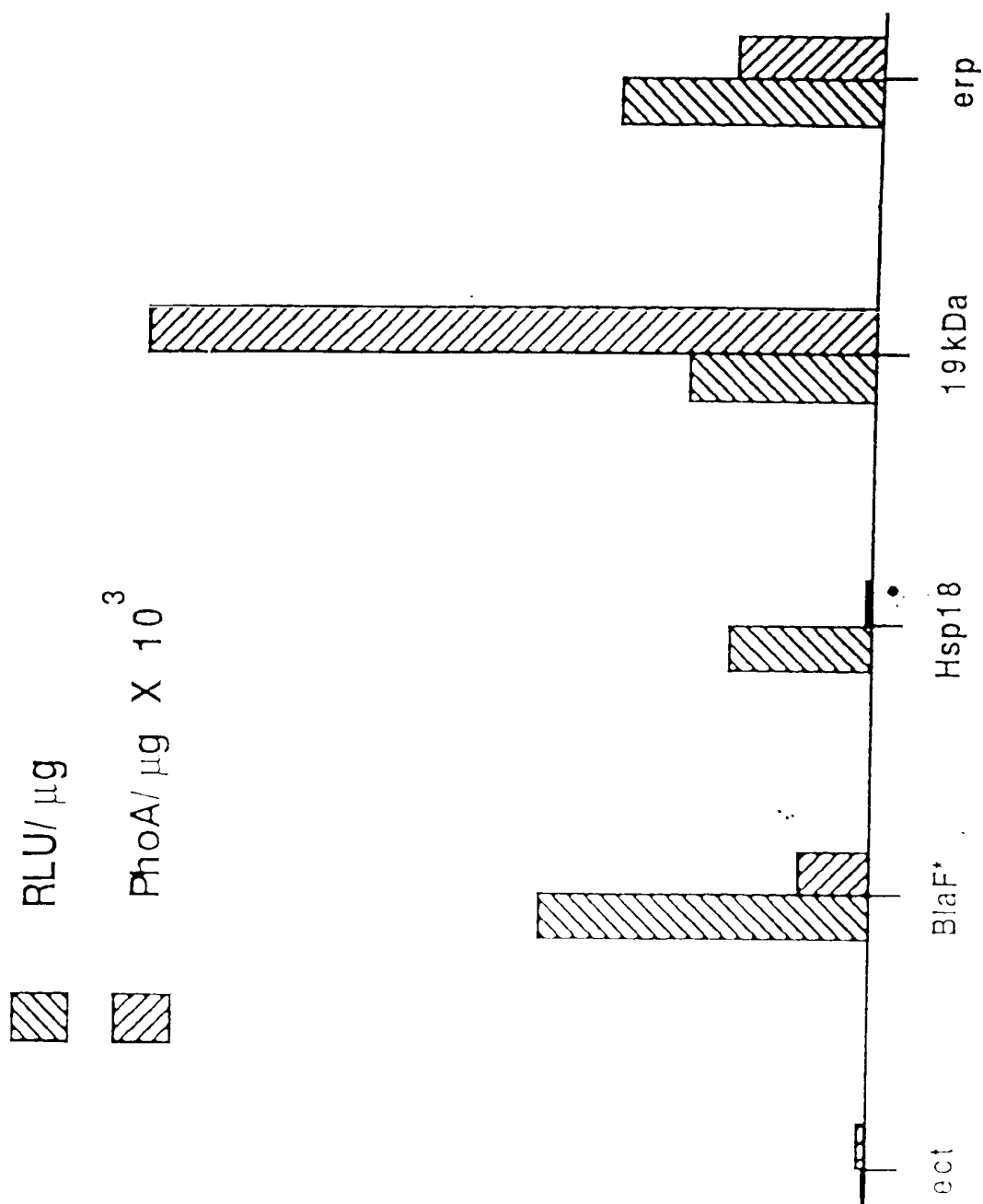


FIGURE 53

184/185

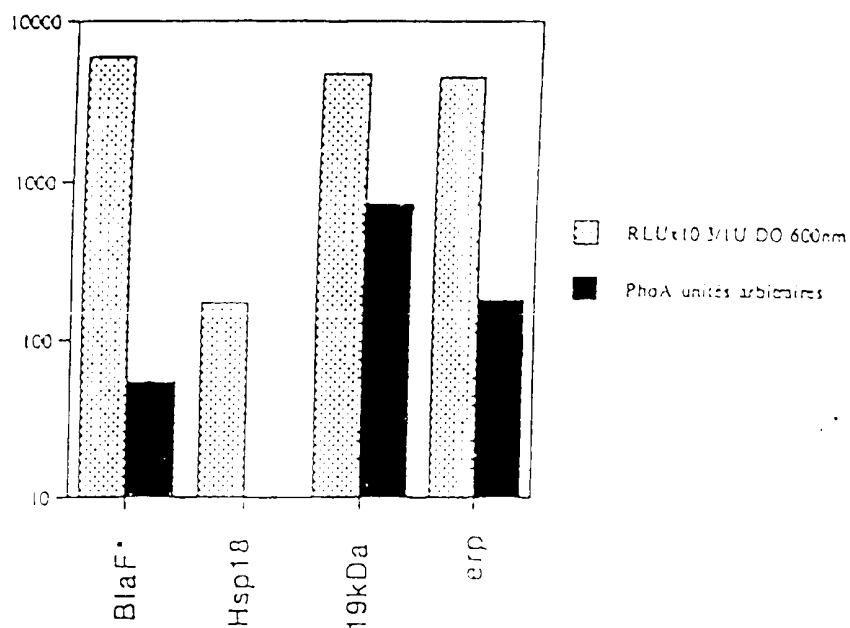
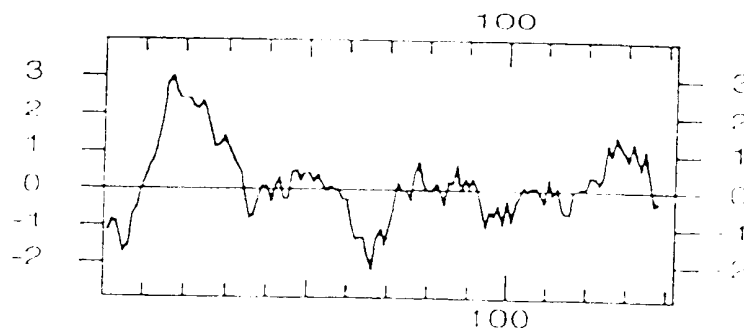


FIGURE 54



signal sequence

0.1

LPISG



185/185

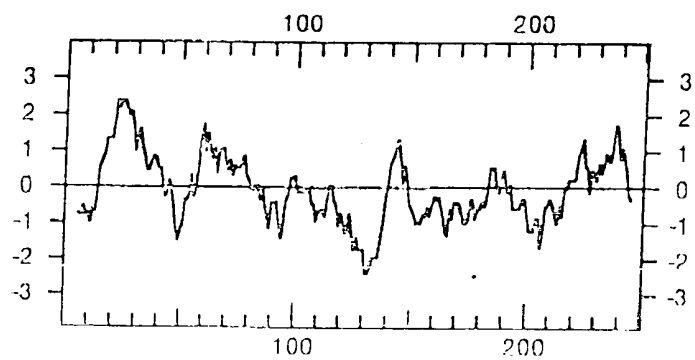


FIGURE 56

